A Publication of the American Psychological Association

The Journal of Abnormal and Social Psychology

Founded by Morton Prince, 1906

CONTENTS

ARTICLES

A Comparison of "Dreamers" and "Nondreamers": Eye Movements, Electroencephalograms, and the Recall of Dreams 295
Donald R. Goodenough, Arthur Shapiro, Melvin Holden, and Leonard Steinschriber

Dominance as a Factor in Interpersonal Choice and Perception 303
John Altrocchi

Stereotyping and Favorableness in the Perception of Negro Faces 309
Paul F. Secord

Judgment of Adjustment by Psychologists, Psychiatric Social Workers, and College Students, and its Relationship to Social Desirability 315
Morton Wiener, Albert Blumberg, Sarah Segman, and Allan Cooper

Some Personality Determinants of the Effects of Participation 322
Victor H. Vroom

Familial Attitudes in Paranoid Schizophrenics and Normals from Two Socioeconomic Classes 328
Robert C. Lane and Jerome L. Singer

Political Orientation and the Perceptual Effects of an Anticipated Interaction 340
Edward E. Jones and Joyce N. Daugherty

Continued on inside front cover

This is the Last Issue of Volume 59
Title Page and Contents for the Volume
Appear Herein.

Volume 59 NOVEMBER 1959 Number 3

CONTENTS—*Continued*

Masculinity, Identification, and Father-Son Relationships 350
Paul Mussen and Luther Distler

Minority Group Belonging, Social Preference, and the Marginal Personality .. 357
D. W. Levit

Reactions to Unfavorable Personal Evaluations as a Function of the Evaluator's Perceived Adjustment 363
Edward E. Jones, Stephen L. Hester, Amerigo Farina, and Keith E. Davis

Test of a Model for Opinion Change 371
Norman H. Anderson

Some Effects of Individually Prominent Behavior upon Group Effectiveness and Member Satisfaction 382
Marvin E. Shaw

Morpheme Order and Syllable Structure in the Learning of Miniature Linguistic Systems 387
Arnold E. Horowitz and H. Merrill Jackson

An Empirical Test of Structural Balance in Sociometric Triads 393
Stephen H. Davol

Extraversion, Neuroticism, and Motor Learning 399
A. W. Bendig and Charles J. Vaughan

Some Relationships between Manifest Anxiety, Authoritarian Tendencies, and Modes of Reaction to Frustration 404
Robert D. Singer and Seymour Feshbach

Discrimination and Reversal Learning in Imbeciles 409
N. O'Connor and B. Hermelin

Identity Diffusion in Late Adolescents 414
Gordon W. Bronson

Accuracy of Brightness Discrimination as Measured by Concurrent Verbal Responses and GSRs 418
Don E. Dulaney, Jr. and Charles W. Eriksen

CRITIQUE AND NOTES

A Search for Correlates of Intra-Individual Response Variability 424
Walter L. Raine and John R. Hills

The Development of Staring Behavior 425
Britomar J. Hamilton and Patricia Gross

Self-Disclosure and Other-Cathexis 428
Sidney M. Jourard

Relationship between Real Similarity and Assumed Similarity with Favorability Controlled 431
David A. Rodgers

Social Desirability and the Description of Others 434
Allen L. Edwards

Value and the Accentuation of Judged Differences: A Confirmation 436
H. Tajfel and S. D. Cawasjee

The Influence of an Affective Set on Disyllable Recognition Thresholds 439
Robert O. Hatfield

Method in Phonetic Symbolism Experiments 441
Roger Brown and Ronald Nuttall

Volume 59 NOVEMBER 1959 Number 3

The Journal of Abnormal and Social Psychology

Editor: M. BREWSTER SMITH, *University of California, Berkeley*

Associate Editor: E. J. SHOBEN, JR., *Teachers College, Columbia University*

CONSULTING EDITORS

GORDON W. ALLPORT
Harvard University

URIE BRONFENBRENNER
Cornell University

ROGER BROWN
Massachusetts Institute of Technology

DONALD T. CAMPBELL
Northwestern University

LAUNOR F. CARTER
System Development Corporation

RICHARD CHRISTIE
Columbia University

MORTON DEUTSCH
Bell Telephone Laboratories

I. E. FARBER
State University of Iowa

DONALD W. FISKE
University of Chicago

ROGER W. HEYNS
University of Michigan

ROBERT R. HOLT
New York University

IRVING L. JANIS
Yale University

HAROLD E. JONES
University of California

HAROLD H. KELLEY
University of Minnesota

LEONARD S. KOGAN
Community Service Society of New York

DANIEL J. LEVINSON
Harvard Medical School

DAVID C. MCCLELLAND
Harvard University

LEO POSTMAN
University of California

ELIOT H. RODNICK
Duke University

NEVITT SANFORD
University of California

SEYMOUR SARASON
Yale University

RICHARD L. SOLOMON
Harvard University

PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

The JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY is devoted to both abnormal and social psychology. Its emphasis is on basic research and theory rather than the techniques and arts of practice. Abnormal psychology is broadly defined to include papers contributing to fundamental knowledge of the pathology, dynamics, and development of personality or individual behavior, including deterioration with age and disease. Articles concerned with psychodiagnostic techniques are evaluated with respect to their contribution to an understanding of the psychological principles of diagnosis, those concerned with psychotherapy are judged in terms of their contribution to an understanding of the therapeutic process. Case reports are considered in terms of their heuristic value. Reports which serve to pose or to clarify important theoretical problems, or which promise to be useful in teaching, are often published. From the social area this JOURNAL gives preference to papers contributing to basic knowledge of interpersonal relations, and of group influences on the pathology, dynamics, and development of individual behavior.

The JOURNAL is issued bimonthly in January, March, May, July, September, and November, two volumes per year. The two volumes average 832 pages. The subscription price per year is $16.00; foreign, $16.50. Single issues are $3.00.

Original manuscripts should be sent to the Editor:

M. BREWSTER SMITH
Department of Psychology
University of California
Berkeley 4, California

ARTHUR C. HOFFMAN
Managing Editor

HELEN ORR
Promotion Manager

BARBARA CUMMINGS
Editorial Assistant

Subscriptions, address changes, and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1333 Sixteenth Street, N.W.
Washington 6, D. C.

Changes of address must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
Mt. Royal and Guilford Aves., Baltimore 2, Md.
and 1333 Sixteenth Street N.W., Washington 6, D.C.

Second-class postage paid at Baltimore, Maryland, and at additional mailing offices

Copyright 1959 by the American Psychological Association, Inc.

A COMPARISON OF "DREAMERS" AND "NONDREAMERS": EYE MOVEMENTS, ELECTROENCEPHALOGRAMS, AND THE RECALL OF DREAMS[1]

DONALD R. GOODENOUGH, ARTHUR SHAPIRO, MELVIN HOLDEN, AND LEONARD STEINSCHRIBER

Downstate Medical Center, State University of New York

THE experimental study of dream phenomena has been seriously handicapped by the lack of effective methods for determining when a sleeping person is dreaming. Recently, however, on the basis of an extensive series of experiments (Aserinsky & Kleitman: 1953, 1955; Dement, 1955; Dement & Kleitman: 1955, 1957a, 1957b), Dement and Kleitman (1957b) concluded that this difficulty can be overcome by studying eye movements and brain waves during the course of sleep.

The major findings of these studies were as follows:

1. There are occasions during sleep when the usual slow drift of the eyes is replaced by rapid conjugate eye movements grossly indistinguishable from the type of eye movements a subject would have, for example, while awake and watching a play.

2. These eye movements occur during light sleep as defined by electroencephalographic criteria (B or A- stages [Simon & Emmons, 1956]).

3. Periods of light sleep during which eye movements occur ranged from 3 to 50 min. in duration and tended to occur periodically throughout the course of natural sleep at intervals of 70 to 104 min.

4. In about 80% of experimental awakenings from eye movement periods, *S*s reported that they had been dreaming, whereas such reports occurred in only about 7% of awakenings at other times. Furthermore, all *S*s studied were found to have eye movement periods and were able to report the occurrence of dreams following most of the eye movement awakenings.

5. The duration of an eye movement period preceding an experimental awakening was consistent with subjective estimates of the duration of the dream.

6. The predominant direction of eye movements (i.e. vertical or horizontal) was related to the direction of movement in the visual imagery of the dream as recalled by the *S*.

These findings are consistent with the proposition that dreams occur during eye movement periods.

If this hypothesis is correct, Dement and Kleitman have discovered a powerful tool for the study of dreams, dreamers, and the process of dreaming. The difference between people who claim that they dream frequently and people who claim that they rarely or never dream is one of the many problems for which this tool provides a method of study.

On the basis of clinical material, observations of sleep talkers, and early studies of the recall of dreams following experimental awakenings (Ramsey, 1953), it has long been clear that many more dreams occur than are usually recalled the morning after. It often has been assumed that the difference between dreamers and nondreamers can be attributed to differences in the recall of dreams rather than to differences in frequency of dream occurrence. This study was designed to check on these hypotheses using the eye movement criteria developed by Dement and Kleitman. More specifically, habitual "dreamers" were compared with habitual "nondreamers" with respect to several characteristics of the eye movement periods as well as the ability to recall dreams following eye movement awakenings and awakenings from periods of ocular quiescence. The study also provides an opportunity to repeat, in an independent laboratory, some of the observations of Dement and Kleitman, on types of *S*s who should offer the most stringent test of the usefulness of the eye movement criteria.

[1] The work was conducted under the auspices of the Psychology Laboratory, Department of Psychiatry. A. Shapiro is a member of the Department of Medicine.

We are grateful to H. A. Witkin for his encouragement and advice and to Michael Tobin, who was primarily responsible for the design and construction of the noncommercial equipment used in this study.

METHOD

Each of the 60 male college students, who were willing to participate in these studies as paid volunteers, filled out a questionnaire indicating how often they usually recalled dreams. From this number, eight *S*s who stated that they recalled dreams almost every night (dreamers) and eight *S*s who stated that they recalled dreams less than once a month (nondreamers) were selected for the laboratory study.

Each of the 16 *S*s slept for three nights in the laboratory (not necessarily consecutively). Each *S* was scheduled to report to the laboratory one hour before his usual bedtime and was awakened for the last time in the morning at his usual hour for rising. The *S* was prepared and instructed during the hour before bedtime. He was told that he would be awakened by a bell; that he could turn off the bell by lifting a bedside phone from its cradle; and that he should report immediately whether or not he had been dreaming and the detailed content of any dream recalled. After practice in the procedure, the *S* was allowed to retire. Electro-oculograms and monopolar frontal, parietal and occipital EEGs were continuously recorded from the time the *S* first went to bed until the final awakening.

The procedures used by Kleitman and Dement (1955, 1957a, 1957b) were followed with few exceptions. Small silver cup electrodes filled with Cambridge Electrode Jelly were attached to the skin by applying U.S.P. collodion and drying with an air jet. Standard frontal, parietal and occipital EEG placements were used. In addition, electrodes were attached to each ear lobe to provide a neutral terminal, to the vertex, and to a point one cm. lateral to the outer canthus of each eye. The individual lead wires were tied together several inches above the vertex and then passed as a single cable to an electrode plug board lying on the floor alongside the bed. A shielded cable led from the plug board through the wall of the bedroom to an eight channel Grass EEG in an adjacent room in which the experimental equipment and *E* were stationed. The *S* was grounded through the shield of the cable and the vertex electrode. Although the bedroom was not soundproof, an attempt was made to keep extraneous noise to a minimum. This attempt was only partially successful.

A recording of a loud bell played through a speaker in the bedroom was used as an awakening stimulus. When the *S* awoke, he answered the bedside phone, automatically terminated the stimulus, and gave his report. *S* could be questioned through an intercom system if the report was not sufficiently clear, but *E* tried not to communicate with *S* after an experimental awakening until the spontaneous report had been completed. All conversation was tape recorded.

An attempt was made to awaken each *S* during at least two eye movement periods and two periods of ocular quiescence per night in counterbalanced order. Two additional awakenings were permitted at the discretion of *E*. However, it was not possible to follow these procedures at all times. In general, the second eye movement period of each night was not interrupted by an experimental awakening.

Nearly all eye movement periods were allowed to continue for at least 5 min., but not longer than 15 min., before an experimental awakening. Awakenings from periods of ocular quiescence always occurred at least 4 min. after the end of any preceding eye movement period.

Criteria for identifying eye movements were identical to those used by Dement and Kleitman. For the purposes of this study, the duration of an eye movement period was measured from the beginning of the first eye movement of the period of light sleep to the last eye movement of the period or to the experimental awakening.

EEG records were inspected and the stage of sleep predominating in each 10-minute period was charted. The times at which each eye movement period began and ended were also recorded. A separate index card was made out for each awakening and for each uninterrupted eye movement period, and the relevant data from the EEG record were noted on one side of the card. All data were rechecked twice and incomplete or incorrectly conducted awakenings were discarded before listening to the tape recordings of the *S*'s report.

The tape recordings were transcribed on the back of the appropriate cards. A "blind" classification of the dream reports was then conducted in terms of the following categories:

D—A dream was recalled in some detail.

ND—The *S* believed that he was definitely dreaming, but could recall no content.

A—The *S* believed that he was awake.

N—The *S* believed that he had been asleep but not dreaming.

The final classification of each report was determined by agreement between two judges. For purposes of subsequent analyses, reports in the D category were taken as evidence of the recall of a dream. Reports classified as ND, A, or N were counted as no dream recall.

The percentage of the awakenings which led to the report of a dream was computed for each *S* under the various conditions of the experiment. All means (and tests of significance) to be reported are based on the percentage per *S* as a descriptive unit. For example, an *S* might recall a dream following 9 out of 10 eye movement awakenings. The *S*'s "score" under these conditions would then be 90%. The mean percentage of recall would be the mean of these scores for all *S*s in the group.[2]

RESULTS

Distribution of Eye Movement Periods

The eye movement periods were distributed throughout the night for both dreamers and

[2] Dement and Kleitman report their data in terms of percentages, summing over all *S*s before dividing the total *N* of awakenings leading to the report of a dream by the total *N* of awakenings. Since the total *N* of awakenings varies from *S* to *S* in these studies the two methods need not lead to similar figures, but in our data they are almost identical.

TABLE 1
RECALL OF DREAMS AS A FUNCTION OF TYPE OF DREAMER AND OCULAR ACTIVITY PRECEDING AWAKENING

Type of Awakening	Dreamers			Nondreamers		
	Total *N* of Awakenings	*N* of Awakenings Leading to Report of a Dream	% Recall of Dreams (Mean per *S*)	Total *N* of Awakenings	*N* of Awakenings Leading to Report of a Dream	% Recall of Dreams (Mean per *S*)
From eye movement periods	49	44	93	42	19	46
From periods of ocular quiescence	56	27	53	43	7	17

nondreamers in the same general pattern described by Dement and Kleitman. Every *S* had at least one eye movement period every night with one exception. (One *S* slept only one-half hour during one of his nights and had no eye movement period that night.) Furthermore, dreamer and nondreamer groups did not differ significantly in the mean number of eye movement periods per hour asleep (Mann-Whitney *U* test[3]). The mean number was .6 per hour for both groups, or one eye movement period every 100 min. of time asleep. (Time asleep was measured from the first appearance of sleep spindles in the EEG record to the final awakening.)

Dream Recall

Every *S*, even if he claimed that he had never dreamt before, was able to report at least one dream during the course of the study. In most cases these reports were given in great detail.

For the entire group of 16 *S*s there were a total of 211 awakenings. However, 21 of these were discarded because of accidents in procedure. Of the remaining 190 awakenings, a mean of 70% of those during eye movement periods and 35% of those during ocular quiescence led to the report of a dream (Table 1). This difference was highly significant ($P < .001$, one-tail sign test).

The difference between awakenings from eye movement periods and awakenings from periods of ocular quiescence was large and significant within the dreamer group ($P < .01$, one-tail sign test) and the nondreamer group as well ($P < .05$, one-tail sign test). All but one *S*, a nondreamer, had a greater percentage of recall following awakenings from eye movement periods.

The dreamers had a significantly higher percentage of dream recall for both types of awakenings than the nondreamers ($P < .02$ and .05, respectively, Mann-Whitney *U* test) (Table 1). The dreamers had a surprisingly high percentage of recall following awakenings from periods of ocular quiescence, but this fact does not necessarily mean that the eye movement criteria are invalid (see discussion).

In order to check on the possibility that these between-group differences might be a function of differences in time asleep, data on the duration of the sleep periods were examined. Table 2 shows that the dreamers tended to sleep longer per night than nondreamers. This difference was due to the fact that the dreamers usually fell asleep earlier.

In order to determine whether the between-group differences in time asleep might account for the differences in dream recall, all awakenings during the last 58 min. of sleep each night were discarded for the *S*s of the dreamer group and the data reanalyzed. This operation equalized the mean duration of the sleep period for the two groups. In all, 29 awakenings were affected, 9 eye movement awakenings and 20 awakenings from periods of ocular quiescence. The mean percentage of eye movement awakenings leading to the recall of a dream was not affected by the correction. The significance of the between-group differences was unchanged. However, for the awakenings from periods of ocular quiescence the mean percentage leading to the recall of a dream fell from 53 to 34. The difference

[3] Reference to the tests of significance used in this study can be found in Siegel (1956).

TABLE 2
DURATION OF THE SLEEP PERIOD AND RELATED DATA

Subjects	Duration of Sleep Period Mean/Night		Time to Bed Mean	Time First Asleep Mean
	Hrs.	Min.	Hr. A.M.	Hr. A.M.
Dreamers	6	:08	12:53	1:08
Nondreamers	5	:10	1:26	2:02
Difference		:58*	:33**	:54*

* $P < .05$ (Mann-Whitney U Test).
** Not statistically significant.

between dreamers and nondreamers in dream recall following such awakenings was no longer significant.

The sharp drop in percentage of recall by dreamers following awakenings from periods of ocular quiescence, obtained by discarding data from the last 58 min. of each night of sleep, suggested that dream recall might be less likely in the early hours of sleep. Awakenings during the first four hours of sleep were compared with awakenings after the end of the fourth hour to test this hypothesis. This division of the sleep period resulted in an approximately equal number of awakenings in the two time intervals.

For awakenings from periods of ocular quiescence in the nondreamer group and eye movement awakenings in the dreamer group, too few responses differed from the prevailing trend to warrant statistical analysis. In the dreamer group, where a mean of 53% of the awakenings from periods of ocular quiescence led to the recall of a dream, an analysis was carried out. Six awakenings which followed an eye movement period by less than 10 min. were not included in this analysis since previous work had suggested that dream recall is frequent following such awakenings (Dement & Kleitman, 1957b). For periods of ocular quiescence, a mean of 17% of the awakenings in the early hours of sleep and 67% of the awakenings in the latter part of the sleep period led to dream recall ($P < .05$, two-tail sign test). A similar analysis was carried out for the eye movement awakenings of nondreamers where 46% led to the recall of a dream. No significant difference in dream recall between early and late sleep periods was found.

Dream recall following eye movement awakenings does not seem to be a function of the time at which the awakening occurs, but dream recall following awakenings from periods of ocular quiescence is more frequent in the last half of the sleep period.

EEG Patterns During Eye Movement Periods

Eye movement periods were classified according to the prevailing EEG pattern during which they occurred. The sleep stages described by Simon and Emmons (1956) were used for this classification.

No rapid conjugate eye movements were observed during EEGs representing deep stages of sleep (C, D, E), although a few low-voltage sleep spindles occurred in some of the eye movement periods. Eye movement periods did occur during the light sleep EEGs which could be classified as A- or B. In order to describe more precisely the EEG records during light sleep, the following procedure was adopted: (*a*) Each 20-sec. interval during an eye movement period was inspected; (*b*) The interval was given an A- classification if any alpha bursts were seen; (*c*) The prevailing sleep stage was taken as A- if more than 50% of these intervals were placed in the A- class. In almost all of the eye movement periods more than 75% or less than 25% of the 20-sec. intervals contained A- bursts. The method of classification thus seemed to reveal a major division in the data.

There seemed to be two distinct types of people among the *S*s with regard to the stage of sleep during which eye movements occurred. For some *S*s all or most of their eye movements occurred during the A- stage. For other *S*s eye movements rarely or never occurred during the A- stage. Six of the nondreamers, but only one of the dreamers, belonged to the A- type. The difference between dreamers and nondreamers was significant ($P < .05$, Fisher's exact probability test).

DISCUSSION

Our results clearly confirm the basic findings of Dement and Kleitman that rapid conjugate eye movements occur during periods of electroencephalographically light sleep and that an *S* who is awakened during such a period is likely to recall a dream. However, the data do not conclusively establish that dreams are

experienced during and only during eye movement periods.

The Eye Movement Definition of Dreaming

The data are consistent with the proposition that a dream is experienced during every eye movement period. It is certainly easier to account for the occurrence of rapid conjugate eye movements behind the closed lids of a sleeping person if it is assumed that a dream is being experienced. Under this assumption the observed differences in dream recall between eye movement awakenings and awakenings from periods of ocular quiescence would also be expected on the basis of well-known principles of forgetting (i.e., the shorter the time interval between an experience and a test of recall, the greater the likelihood of recall).

The finding that there are some eye movement awakenings which do not lead to the recall of a dream is important, but not necessarily inconsistent with the assumption. In fact, some of the most interesting implications of the eye movement definition of dreaming have to do with the forgetting of dreams. In order to identify instances of dream forgetting, it is necessary to know when a dream has taken place. The eye movement criteria provide a working definition of dreaming with which the study of forgetting may be approached.

The proposition that dreams occur only during eye movement periods is more difficult to justify on the basis of the available evidence.

Periods of light sleep without conjugate eye movements sometimes occur, and Dement and Kleitman (1957a) have suggested that dreams are experienced during these periods. In this study there were too few awakenings from light sleep without eye movements to test their suggestion.

The large number of dream reports given after awakenings from deep sleep poses a more serious problem. It is possible to account for these reports, however, without abandoning the assumption that dreams occur only during eye movement periods or light sleep.

It is suggested that dream reports following awakenings from periods of ocular quiescence may refer to dreams which occurred earlier in the night during eye movement periods, or even to the recall of thoughts which the *S* might have had while falling asleep. This interpretation is based on two reasonable assumptions:

1. Dement and Kleitman (1957b) found that dreams are often recalled following awakenings which occur less than eight minutes after the end of an uninterrupted eye movement period. These reports presumably reflect recall of dreams which occurred during the preceding eye movement period. It is assumed that there may be some recall of dreams over much longer periods of time, particularly for the "dreamers" of this study.

2. If an *S* is awakened and recalls a dream, it is assumed that he may have difficulty determining whether the dream occurred just before the awakening or at an earlier time.

Such an interpretation seems reasonable on the basis of qualitative observations made in the course of this study. Several of the dream reports given after awakenings from periods of ocular quiescence were very similar in content to earlier reports by the same *S*s on the same night following eye movement awakenings. In one case, the *S* spontaneously said that he wasn't sure whether he was repeating an earlier report or describing a dream that had just occurred.

The data on the percentage of awakenings from periods of ocular quiescence which led to the report of a dream are also consistent with this interpretation. Since awakenings late in the night are necessarily preceded by more dreaming than early awakenings, the finding that dream recall is more frequent in the late part of the sleep period is to be expected.

Under the assumption that the process of recall is impaired in deep sleep, it could be argued that dreams are occurring even though no dream is ever reported following an awakening from such deep sleep. If it is also assumed that eye movements are blocked in deep sleep, it seems possible to account for the findings of these studies. However, we believe that it will prove more fruitful to use the eye movement criteria as a working definition of the dreaming state. The use of this definition apparently leads to interesting and testable hypotheses.

The Forgetting of Dreams

If it is assumed that dreams are experienced during and only during eye movement periods (or light sleep) our data indicate that the difference between habitual dreamers and habitual nondreamers does not lie in the frequency of dreaming. The groups do differ in the ability to recall dreams even under the conditions of this experiment.[4]

It has often been suggested that massive repression plays an important role in the forgetting of dreams (Freud, 1956). Of course our data do not bear upon this hypothesis, but they suggest a method for future research. Since eye movement periods which lead to the recall of a dream can be distinguished from those which do not, it is possible to study the conditions under which forgetting occurs. It is also possible to study dream content among *S*s who do not ordinarily recall their dreams. Such studies may provide a new and useful method for systematic experimental analysis of the mechanism of repression.

Our data suggest another hypothesis that might help to account for failure to report dreams under nonlaboratory conditions.

If a dream is defined as an experience that occurs during sleep, some *S*s probably mislabel their experiences. One *S* (from a different study) said that he was ***asleep*** and ***thinking*** when awakened from an eye movement period. In other instances, *S*s were not sure whether they had been asleep and dreaming or awake and thinking. Neither type of mislabelling is very frequent under the conditions of this experiment, but the occurrence of any mislabelling raises some interesting questions.

During the instruction period one of the *S*s who said that he never dreamt asked this question: "How will I know if I have had a dream or not?" Although we were somewhat surprised by it at the time, this seems like a reasonable question from a person who has never recalled a dream.

What cues does an *S* use to judge whether or not he has had a dream?

The popular concept of a dream certainly includes the notion that it is an experience which occurs during sleep. It might also include the following characteristics: (*a*) The dream is experienced as if something were happening to the dreamer as opposed to an imagined series of events; (*b*) It is not accompanied by the feeling of control over the content; (*c*) It is frequently bizarre or internally inconsistent.

The *S* who said that he was asleep and thinking, apparently felt that he had had an experience during sleep which was not bizarre or internally inconsistent and over which he felt some control, etc. In fact, a large number of the dream reports collected in this study were not bizarre or internally inconsistent, even though they were given in some detail. We doubt that they could be distinguished from reports of some of the hypnogogic reveries or waking fantasies of these *S*s. However, most *S*s would probably have said that they were dreaming if they had known (or believed) that they were asleep at the time, even if the "typical" content characteristics were not present.

Let us suppose that an *S* recalls an experience when he awakens in the morning under ordinary conditions. How does he decide whether he was asleep or awake at the time the experience occurred?

In the absence of any other, more decisive, cues, it seems likely that the content of the recalled experience may be used by an *S* to judge whether he was asleep and dreaming or awake and thinking at the time. For example, if an *S* recalls an experience that is bizarre or internally inconsistent he might decide, on this basis alone, that he was asleep and dreaming. The following verbatim reports are illustrative:

> I am not sure whether I fell asleep. Oh, yes, I was at the beach. I was with my mother and father and a girl next door. She is a friend of the family, not too much a friend of mine. We were talking, and I couldn't sit next to her, so I was somewhat displeased, so I changed to another seat. I started talking to other girls and so did she. I can remember specific conversation. People came up and asked me why I was carrying my bag—clothing bag. I looked in a newspaper for a certain basketball score. Basketball is in the winter. I think I must have been asleep. I wouldn't have been thinking about this if I were awake, I don't imagine. I guess I was dreaming.

[4] This difference may be reduced through the use of different awakening techniques, however. In more recent studies we have obtained much higher percentage of recall from nondreamers by using improved sound conditioning and a louder awakening stimulus.

From the *S* who had asked how he could tell whether or not he was dreaming:

> I don't know whether I was dreaming or not. I'm not sure whether I was asleep or not. I had a vision in mind of riding down a wide avenue—no story, but moving, as if sitting in an automobile and the various houses and buildings passing by—a vision. I may have thought that or dreamt it.

In the first illustration, the *S* apparently decided that he was asleep and dreaming on the basis of an internal inconsistency in the content. In the second illustration, no inconsistency appears in the report and the experience is not definitely identified as a dream.

The finding that eye movement periods tend to occur in association with A- EEGs for nondreamers, and B EEGs for dreamers is particularly interesting in the light of the mislabelling hypothesis. If this finding is confirmed by future research, it may clarify the basis for the failure to report dreams in some *S*s.

According to the traditional view (Simon & Emmons, 1956) an A- EEG represents a borderline state between drowsiness and sleep. Following this view it might be argued that for many of our eye movement awakenings the *S* was not actually asleep. Kleitman (1957) discounts this argument as follows:

> Just as a "sleep" EEG is without diagnostic significance in the presence of behavioral wakefulness, ... so is a "wakefulness" EEG, obtained during behavioral sleep. Such an EEG was invariably observed in our laboratory during dreaming, with the *S*s unquestionably asleep (p. 357).

While our study was not designed to check on this question, systematically, our impressions are the same as Kleitman's. The errors in subjective judgment cited in the illustrations should not be taken as evidence against Kleitman's contention. These errors occurred with as large a relative frequency in deep sleep as in light sleep. In all, 17% of the awakenings led to the judgement, "I was awake."

An A- EEG record may reflect a lighter stage of sleep than a B EEG record, however. Furthermore, it seems reasonable to suppose that dreams which occur during a very light stage of sleep would not be as bizarre and would seem more like thought or fantasy to the *S* than would dreams which occur in a deeper stage of sleep. In this view the traditional distinction between the kind of thought processes that usually occur during the waking state and the kind that usually occur during dreams is expanded. Three stages, defined in terms of depth of sleep (or state of cortical functioning) and type of thought processes are proposed.

If the dreams that occur during an A- EEG are less bizarre, etc., and if such content characteristics are important determinants of recall or correct labelling under ordinary conditions, then it is not surprising to find some nondreamers who have most of their eye movement periods during A- EEGs.

There is another interpretation of this EEG difference that may have some merit.[5] Grey-Walter (1953) has identified individual differences in waking EEG records, related to differences in characteristic thought processes. In most adult *S*s alpha waves can be observed in waking occipital tracings recorded under relaxed conditions with eyes closed. These waves are usually blocked for a period of time following the sudden presentation of a visual stimulus. The resulting EEG record is similar to (if not identical with) the record obtained during the B stage of sleep. However, in some *S*s alpha blocking does not occur. For these *S*s, the alpha waves persist despite a sudden visual stimulus. Grey-Walter believes that these *S*s tend to think in symbolic terms with little visual imagery. Other *S*s have little or no alpha under any conditions. These *S*s are believed to use visual imagery extensively, with little recourse to symbolic thought. The majority of people, who show a waking alpha that can be blocked by a visual stimulus, presumably use visual imagery or symbolic thought as the situation requires.

Only one of our *S*s had no waking alpha, but, unfortunately, studies of alpha blocking were not conducted. It may be that *S*s who have their eye movement periods during the A- sleep stage also show a waking alpha that is resistant to blocking, and characteristically think in symbolic terms, whether awake or asleep. Such continuity between the sleeping

[5] We are indebted to W. Dement for calling our attention to this possibility.

and waking states seems reasonable enough on a priori grounds.

Although these hypotheses have been based in part on physiological concepts, personality characteristics of the dreamer may play a crucial role. It is assumed that dreams which occur during A- EEGs are accompanied by some feeling of control over the content. If these hypotheses are confirmed, it would be interesting to explore the possibility that A- dreams may occur in *S*s who are overcontrolled more generally in their psychological functioning.

SUMMARY

Habitual dreamers were compared with habitual nondreamers using Dement and Kleitman's "objective method for the study of dreaming." Two groups of *S*s were studied, a group of dreamers who reported that they usually dreamt almost every night and a group of nondreamers who said that they dreamt less often than once a month. Continuous recordings of eye movement potentials and EEGs were taken on each *S* during three nights of natural sleep. *S*s were awakened several times each night by the sound of a loud bell and were asked whether or not they had been dreaming.

Periods of rapid conjugate eye movements occurred at intervals throughout the night during light sleep (A- or B stages) for all *S*s. Experimental awakenings from eye movement periods led to the recall of dreams more frequently than awakenings at other times. So-called "nondreamers" were less likely to recall a dream than dreamers, but every *S* studied, even *S*s who said that they had never dreamt before, reported at least one dream during the study.

Eye movement periods occurred as frequently for nondreamers as for dreamers. However, there were significant differences between dreamers and nondreamers in the EEG patterns which occurred during eye movement periods.

These findings support the contention of Dement and Kleitman (1957b) that recordings of EEG and eye movement potentials may provide useful objective criteria for determining when a dream is occurring. They also suggest directions for future research on problems relating to the forgetting of dreams.

REFERENCES

ASERINSKY, E., & KLEITMAN, N. Regularly occurring periods of eye motility, and concomitant phenomena, during sleep. *Science*, 1953, **118**, 273–274.

ASERINSKY, E., & KLEITMAN, N. Two types of ocular motility occurring in sleep. *J. appl. Physiol.*, 1955, **8**, 1–10.

DEMENT, W. Dream recall and eye movements during sleep in schizophrenics and normals. *J. nerv. ment. Dis.*, 1955, **122**, 263–269.

DEMENT, W., & KLEITMAN, N. Incidence of eye motility during sleep in relation to varying EEG pattern. *Fed. Proc.*, 1955, **14**, 37.

DEMENT, W., & KLEITMAN, N. Cyclic variations in EEG during sleep and their relation to eye movements, body motility, and dreaming. *EEG clin. Neurophysiol.*, 1957, **9**, 673–690. (a)

DEMENT, W., & KLEITMAN, N. The relation of eye movements during sleep to dream activity: An objective method for the study of dreaming. *J. exp. Psychol.*, 1957, **53**, 339–346. (b)

FREUD, S. *The interpretation of dreams*. New York: Basic Books, 1956.

GREY-WALTER, W. *The living brain*. New York: Norton, 1953.

KLEITMAN, N. Sleep, wakefulness, and consciousness. *Psychol. Bull.*, 1957, **54**, 354–359.

RAMSEY, G. Studies of dreaming. *Psychol. Bull.*, 1953, **50**, 432–455.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.

SIMON, C. W., & EMMONS, W. H. EEG, consciousness, and sleep. *Science*, 1956, **124**, 1066–1069.

(Early publication received May 18, 1959)

DOMINANCE AS A FACTOR IN INTERPERSONAL CHOICE AND PERCEPTION[1]

JOHN ALTROCCHI

Duke University

TO INVESTIGATE the possibility that personality variables may be factors in whom one likes and how one perceives others, this study focused on two questions: What is the relationship between the degree of dominance in a person and the degree of dominance in persons to whom he is attracted? What is the relationship between the degree of dominance in a person and the degree of dominance he attributes to others?

In interpersonal choice, arguments have been going on for centuries as to whether "birds of a feather flock together" or "opposites attract." Some studies have found low but usually positive correlations between friends and marriage partners in personality variables such as dominance (Corsini, 1956; Lindzey & Urdan, 1954; Richardson, 1939). In contrast, some theories of interpersonal choice state that people are attracted to others with personalities complementary to their own. Leary (1957) suggests that an individual seeks to interact with those whose behavior is reciprocal to his. Consistent with Leary's theory is his observation that dominant and submissive individuals in therapy groups tend to seek each other out. A more specific theory of interpersonal choice is offered by Winch, Ktsanes, and Ktsanes (1954) in their theory of selective interaction, a special case of which is the theory of complementary needs in mate selection. In two studies designed to test the theory, Winch (1955) and his co-workers (Winch, Ktsanes, & Ktsanes, 1955), found some support for the suggestion that people with dominant needs tend to be married to people with submissive needs. Although this theory is subject to several reservations (cf. Altrocchi, 1957; Kirkpatrick, 1952; Rosow, 1957; Winch, 1952), it is used, in combination with Leary's thinking, as a basis for Hypothesis 1 in this study:

Hypothesis 1. People choose to interact with people complementary to themselves in dominance. In Winch's terms, a dominant person tends to choose to interact with a person likely to satisfy his need to dominate—i.e., a submissive person. In Leary's terms, a submissive person tends to behave in the way that a dominant person forces people to behave and therefore the dominant person will choose to interact with him.

Interpersonal perception is defined, in this study, as the attribution of relatively enduring personality traits to other persons. Bruner and Tagiuri (1954) suggest, and Naboisek (1953), Leary (1957), and Smelser (1957) present evidence that a dominant person is more likely to see others with whom he is interacting as submissive and a submissive person is more likely to see others with whom he is interacting as dominant. Therefore:

Hypothesis 2. Dominant people perceive other people as less dominant than submissive people perceive them to be. With the same cues available, a dominant person should be likely to attribute less dominance to others than a submissive person.

METHOD

Subjects (*Ss*) were shown a movie taken of the interaction between four social objects who differed in dominance, and then asked to state their preference for the social objects and to describe them.

Instruments

Gough's California Psychological Inventory (CPI) Dominance (Do) scale (Gough, 1957) was used to provide an estimate of characteristic interpersonal dominance behavior. This 46-item scale was constructed, validated, and cross-validated against ex-

[1] This paper is based on a dissertation submitted in partial fulfillment of the requirements for the Ph.D. degree in psychology at the University of California. The writer wishes to express his appreciation for their help to David A. Rodgers, Harrison Gough, William Smelser, Lyman Porter, Edward Barnhart, Nevitt Sanford, Robert Tryon, Timothy Leary, Harold Sampson, and Read Tuddenham; to Karl Zener who provided financial assistance for added statistical analysis; and to Edward E. Jones and Charles Brodie for critical reading of drafts of this paper. An earlier version of this paper was read at the meetings of the Southeastern Psychological Association, Atlanta, Georgia, April, 1958.

ternal criteria of dominance behavior (Gough, 1957; Gough, McClosky, & Meehl, 1956). The Dominance score (Dom) from the Interpersonal Check List (ICL) (Leary, 1957) was used as a measure of dominance consciously attributed to self and others. Interpersonal choice was measured by the *Ss*' ranking of the social objects on five quasi-sociometric questions. These questions were selected, on the basis of pretesting, to cover a wide range of situations relevant for college males (e.g., "... which one would you like to room with").

Social Objects

Four social objects were selected from a pool of 200 male college students who had been administered the CPI and a vocabulary test (Thorndike & Gallup, 1944). Through interviewing and retesting, four social objects with the following characteristics were chosen: Two could be expected to behave in a moderately dominant fashion, and two could be expected to behave in a moderately submissive fashion (Do *T* scores of about 60 and about 40 respectively); the degree of dominance they attributed to themselves was consistent with their Do scores (ICL Dom *T* scores in the 60's and 40's respectively); all four had similar scores on the vocabulary test, the Responsibility and Self-Control scales of the CPI, and the Love dimension of the ICL; all were white, male sophmores, strangers to each other, who demonstrated no other outstanding physical or personal characteristics on the tests or in the interview.

The Movie

The social objects were dressed identically except for single identifying letters on their caps. These letters (B, D, L, and T) were chosen because they proved, in pretesting, to be of approximately equal preference value. The interaction situation selected to allow them to demonstrate different degrees of dominance behavior was a model railroad problem-solving apparatus (Ghiselli & Lodahl, 1958). In previous experiments, male students had quickly become involved in this task and spontaneously demonstrated different degrees of dominance behavior. Four hundred feet of sound movies were taken of the interaction during parts of a 40-minute period. Cutting resulted in a 10-minute movie with a good picture but only fair sound.

Two separate pretestings of the movie demonstrated that *Ss* generally perceived the expected differences in dominance behavior and, except for Object T who was slightly lower, saw the social objects as approximately equal in likability. It seemed reasonable to assume that differential choice of L, B, D, and T would not be due to general differences in likability and might be at least partially due to their differences in dominance behavior.

Subjects and Procedure

Ninety-six male students who did not know any of the social objects were drawn from the same pool as the social objects. In groups varying in size from 1 to 18 they described themselves, their best male friend, and ideal male friend on the ICL, saw the movie, answered the choice questions, and described L, B, D, and T on the ICL (the order of description being varied in a latin square design).

RESULTS

Choice was measured by the total sum of ranks which an *S* assigned to a social object on the five personal choice questions. Application of Tryon's formulas (1957) suggested that it was justifiable to combine the five questions and that this combined score (called P) was adequately reliable and distributed relatively normally.

Analysis of variance of P scores provided a first test of Hypothesis 1. The *Ss*, divided into four groups (quartiles, except for ties) on Do made up one variable, the objects the other variable. (See Table 1). The interaction variance is not significantly different from zero. Thus no relationship is demonstrated between an *S*'s Do score and whom he chose.[2] The objects variance is significant. This was due to the fact that T was less often chosen. Thus the selection procedure was not entirely successful in selecting social objects who would be chosen equally often.

Hypothesis 1 was also tested by ascertaining if there was a linear relationship between *Ss*' Do scores and the P scores they assigned to each object. These four product-moment correlations were: T, − .11; D, + .11; B, + .12; and L, −.13. Again there is no evidence of a relationship between Do and choice.

For comparison with many of the earlier studies of interpersonal choice, the descriptions of best friend (BF) and ideal friend (IF) were employed. Best friends were described as rather dominant, and ideal friends were, with little variation, described as quite dominant (see Table 3 below). Thus the *Ss* attributed considerable dominance to these esteemed people. This tendency was slightly more marked for more dominant *Ss*, with correlations between the two dominance measures and BF and IF ranging from .17 to .29 ($P < .01$). These correlations are similar to those in earlier studies (cf. Richardson, 1939). These results,

[2] It can be argued that this analysis is not appropriate because the scores in each column are not independent of those in the other columns. Therefore, another analysis of variance was performed, using random numbers to divide the *Ss* in each row into four independent sets of five *Ss* each. In this analysis, too, the interaction variance was not significant.

TABLE 1
ANALYSIS OF VARIANCE FOR DOMINANCE AND CHOICE (P SCORES)

Source	Sum of Squares	*df*	Variance Estimate	*F*	*P*
Between subjects (Do scores)	0.0	95			
Subjects	0.0	3	0.0		
Error (b)	0.0	92	0.0		
Within subjects	8,880.50	288			
Objects	509.91	3	169.97	5.60	<.001
Subjects × objects	478.21	9	53.13	1.75	<.10 >.05
Error (w)	8,370.59	276			
Total	8,880.50	383			

however, may merely reflect a tendency to describe one's friends in socially desirable terms (Edwards, 1957).

Analysis of variance of attributed Dom scores provided a test of Hypothesis 2. To secure independence of scores for entry into the analysis, random numbers were used to divide the *S*s in each group into four independent sets of five *S*s each.[3] (See Table 2). Neither the *S*s variance nor the interaction variance is significantly different from zero. Thus there is no evidence that any of the four groups of *S*s attributed more or less dominance to any object.

The objects variance is highly significant. Thus there was a general tendency for the *S*s to attribute different degrees of dominance to the social objects. (See Table 3). The dominance of the four social objects is in the predicted order, but the goal of having two highs and two lows was not entirely achieved because D was seen as more dominant than expected. The four objects, however, were seen as covering a fairly wide range, within normal limits, on dominance.

Hypothesis 2 was also tested by ascertaining if there was a linear relationship between *S*s' Do scores and the Dom scores they assigned to each object. These four correlations were: T, −.10; D, .00; B, +.04; L, +.04. Again there is no evidence of a relationship. In addition, however, the correlations between Dom attributed to self and Dom attributed to social objects were: T, .01; D, .03; B, .02; L, .10. Thus there is no evidence that the *S*s' descriptions of the objects were dependent on the way the *S*s described themselves.

[3] It can be argued that this analysis wastes data, but the data do not allow a more complex analysis. A three-way analysis (using the latin square order of describing the objects) can not be used without throwing away much of the data, since the cell frequencies in such an analysis vary from 3 to 10.

TABLE 2
ANALYSIS OF VARIANCE FOR DOMINANCE AND PERCEPTION (ICL DOM)

Source	Sum of Squares	*df*	Variance Estimate	*F*	*P*
Subjects (Do Scores)	160.15	3	53.38	.51	—
Objects	4358.45	3	1452.82	13.65	<.001
Subjects × objects	1479.75	9	164.42	1.55	—
Within cells	6809.60	64	106.40		
Total	12807.95	79			

TABLE 3
ATTRIBUTED DOMINANCE SCORES

Person Described	*m*	*SD*
T	45.3	12.2
D	56.0	11.7
B	59.2	10.6
L	66.8	8.4
Self	55.9	8.6
BF	59.8	6.1
IF	65.1	3.8

Further examination of the results (Altrocchi, 1957) revealed that: *S*s described themselves as more dominant than their peers would describe them to be (mean Do minus mean Dom = −4.2), but these two measures of dominance were still highly related (r = .65); and there was a significant ($P < .01$) positive relationship between choosing T and D and attribution of dominance to them (r's of .48 and .54 respectively). Having chosen a less dominant person, an *S* tended to attribute more dominance to him. Since dominance is a positively valued trait in our culture, for males at least, it may be hard to separate it from social desirability (Edwards, 1957) in systems like Leary's.

DISCUSSION

In the light of the clear failure of the results to support the hypotheses, it may be important to examine the method more closely. The

instruments used were satisfactorily reliable. The CPI Do scale predicted the dominance behavior of the social objects fairly accurately, and the ICL Dom scores were consistent with the Do scores for both *S*s and social objects. The secondary results were consistent with expectations (e.g., the positive correlations between choosing T and D and attributing dominance to them). Finally, the results were consistent with previous evidence where similar procedures were used (as in the results for best and ideal friends). Therefore it seems fair to assume that the negative results can not be attributed to unreliable or invalid instruments.

The method of testing the hypotheses was fairly direct. The use of a movie focuses on first impressions of people with whom the *S*s did not feel highly involved, but such first impressions are implicitly included in the theories. While the use of only four social objects limits generalization from the results (Crow, 1957), it does not weaken the test of the hypotheses. Thus it may be concluded that the study was a fairly valid test of the hypotheses. These hypotheses were derived in a straightforward manner from the theories. Therefore it seems appropriate to explore the lack of confirmation of the hypotheses with a view toward revising the theories from which they were derived.

The theory that opposites attract, which Winch and Leary find supported by their data, is not supported by the present results. The fact that their *S*s, however, were married couples and group therapy patients who had interacted together over a period of time and had had a chance to establish reciprocal role relationships suggests an alternative interpretation of their data. Winch's marital partners may have developed complementary need patterns as a *result* of their process of establishing a reciprocal role relationship. Leary's patients may have *seemed* to interact more with persons complementary to themselves, not because of initial attraction, but because the patients influenced one another to react in a complementary or reciprocal fashion. This line of reasoning moves toward viewing interactions as being determined by situational as well as personality factors. If this line of reasoning is valid, it would be necessary to substitute a theory of differential influence (which can be seen as an extension of Leary's [1957] Principle of Reciprocal Interpersonal Relations) for the theories of selective interaction. If this line of reasoning is not valid, it might at least be advisable to exclude brief and uninvolved relationships from the theories, or to make the theories more complex, as suggested by Rosow (1957). For instance, Colvin has presented evidence to support the hypothesis that "opposites attract" when self-acceptance is lacking, and that "birds of a feather flock together" when self-acceptance is present (Colvin, 1957).

The contrast between the present results concerning interpersonal choice and early studies which tended to show that similar personalities are attracted to each other (cf. Richardson, 1939) can be explained fairly directly. In many of these studies the people choosing provided the personality ratings of the persons chosen, so that the results do not necessarily indicate that people attracted to each other actually are similar. The results may only indicate that people tend to describe people they like as similar to themselves (cf. the correlations between self-description and description of best friend and ideal friend in the present study). As Lindzey and Borgatta (1954) have pointed out (p. 433), in order to obtain evidence concerning the actual similarity between persons attracted to each other, it is necessary to utilize externally validated personality measures.

The results concerning interpersonal perception suggest that *S*s tended to attribute more dominance than would be justified to themselves and to submissive social objects whom they liked, but that the degree of dominance attributed to the social objects was not a function of the *S*s' own dominance behavior. The latter results contrast directly with conclusions from Naboisek (1953), Leary (1957), and Smelser (1957). The contrast between evidence from the present study and Smelser's is particularly striking, because the two studies used *S*s from the same pool of students and used identical measures of dominance. The crucial difference may be that Smelser's pairs of *S*s interacted together for an hour, while *S*s and objects in this study never did so. Thus no probing or eliciting of cues was possible.

Bruner and Tagiuri (1954) suggest the relevance of the cue-evoking power of the perceiver: "Many of the cues used in judging another person are cues that we as perceivers are instrumental in producing .... Do dominant individuals, for example, have the effect of evoking noticeably submissive behavior in their less dominant peers, with the result that other people are seen as more passive?" (p. 649). This cue-evoking power of the perceiver, or differential influence, may explain the contrasting results. If a dominant person has an opportunity to interact with others, he instigates submissiveness in them; then his description of them as submissive accurately reflects their behavior with him. If, however, he does *not* have a chance to interact with the others, he will describe them no differently from the way submissive people describe them. If this explanation is correct, then misperception or perceptual distortion are not appropriate terms for what Sullivan (1947), Rogers (1951), and Leary (1957) describe. These theorists may have observed correctly that their patients perceive others differently from the way normals perceive them, but this may be correct perception because the patients consistently influence or train others to behave differently *with them*.

SUMMARY

This study investigated the relationship between the degree of dominance in a person and (*a*) the degree of dominance in persons to whom he is attracted; (*b*) the degree of dominance he attributes to others. A 10-minute sound movie was taken of the interaction between four male social objects who were selected because they were different in dominance but similar in likability. Ninety-six male subjects viewed the movie, ranked the social objects on five-choice questions, and described themselves and the objects on the Interpersonal Check List. Gough's Dominance scale provided an estimate of behavioral dominance.

Whom a subject chose and how much dominance he attributed to the social objects were found to be unrelated to his dominance. These results suggest modifications in interpersonal theory and in conclusions drawn from earlier research. Particularly, they suggest that differential influence of one person on another may affect the phenomena of both interpersonal choice and perception. Thus, when no interaction between subjects and objects occurs, as in this study, a person has no chance to elicit cues from, or influence the behavior of, the other person. When interaction does occur, a person may influence others to behave in ways complementary to his own. Therefore, the people he interacts with appear to be complementary to him and, furthermore, he is justified in perceiving them as complementary to himself, his deviant perceptions of them not necessarily being misperceptions.

REFERENCES

ALTROCCHI, J. C. Dominance as a factor in interpersonal choice and perception. Unpublished doctoral dissertation, Univer. of California, 1957.

BRUNER, J. S., & TAGIURI, R. The perception of people. In G. Lindzey (Ed.), *Handbook of social psychology*. Cambridge, Mass.: Addison-Wesley, 1954. Ch. 17.

COLVIN, R. W. Friendship selections as a defense of the self-concept. Paper read at American Psychological Association meeting, New York, September, 1957.

CORSINI, R. J. Understanding and similarity in marriage. *J. abnorm. soc. Psychol.*, 1956, **52**, 327–332.

CROW, W. J. The need for representative design in studies of interpersonal perception. *J. consult. Psychol.*, 1957, **21**, 323–325.

EDWARDS, A. L. Social desirability and probability of endorsement of items in the Interpersonal Check List. *J. abnorm. soc. Psychol.*, 1957, **55**, 394–396.

GHISELLI, E., & LODAHL, T. M. Patterns of managerial traits and group effectiveness. *J. abnorm. soc. Psychol.*, 1958, **57**, 61–66.

GOUGH, H. G. *Manual for the California Psychological Inventory*. Palo Alto, Calif.: Consulting Psychologists Press, 1957.

GOUGH, H. G., MCCLOSKY, H., & MEEHL, P. E. A personality scale for dominance (Do). In G. S. Welsh & W. G. Dahlstrom (Eds.), *Basic readings on the MMPI in psychology and medicine*. Minneapolis: Univer. Minnesota Press, 1956. Pp. 212–220.

KIRKPATRICK, C. Review of Winch's The Modern Family. *Amer. sociol. Rev.*, 1952, **17**, 375–376.

LEARY, T. *Interpersonal diagnosis of personality*. New York: Ronald, 1957.

LINDZEY, G., & BORGATTA, E. F. Sociometric measurement. In G. Lindzey (Ed.), *Handbook of social psychology*. Cambridge, Mass.: Addison-Wesley, 1954. Ch. 11.

LINDZEY, G., & URDAN, J. A. Personality and social choice. *Sociometry*, 1954, **17**, 47–63.

NABOISEK, H. Validation of a method for predicting role expectations in group therapy. Unpublished doctoral dissertation, Univer. of California, 1953.

RICHARDSON, HELEN M. Studies of mental resemblance

between husbands, wives, and between friends. *Psychol. Bull.*, 1939, **36**, 104–120.

ROGERS, C. R. *Client-centered therapy*. New York: Houghton Mifflin, 1951.

ROSOW, I. Issues in the concept of need-complementarity. *Sociometry*, 1957, **20**, 216–233.

SMELSER, W. T. Personal communication, 1957.

SULLIVAN, H. S. *Conceptions of modern psychiatry*. Washington, D. C.: William Alanson White Psychiatric Found., 1947.

THORNDIKE, R. L., & GALLUP, G. H. Verbal intelligence of the American adult. *J. genet. Psychol.*, 1944, **30**, 75–85.

TRYON, R. C. Reliability and behavior domain validity: Reformulation and historical critique. *Psychol. Bull.*, 1957, **54**, 229–249.

WINCH, R. F. *The modern family*. New York: Holt, 1952.

WINCH, R. F. The theory of complementary needs in mate-selection: A test of one kind of complementariness. *Amer. sociol. Rev.*, 1955, **20**, 52–56.

WINCH, R. F., KTSANES, T. & KTSANES, V. The theory of complementary needs in mate selection: An analytic and descriptive study. *Amer. sociol. Rev.*, 1954, **19**, 241–249.

WINCH, R. F., KTSANES, V., & KTSANES, T. Empirical elaboration of the theory of complementary needs in mate selection. *J. abnorm. soc. Psychol.*, 1955, **51**, 508–513.

(Received July 18, 1958)

STEREOTYPING AND FAVORABLENESS IN THE PERCEPTION OF NEGRO FACES[1]

PAUL F. SECORD
University of Nevada

A STEREOTYPE is commonly thought of as involving a categorical response; i.e., membership in a category is sufficient to evoke the judgment that the stimulus person possesses all the attributes belonging to that category. In a previous study, Secord, Bevan, and Katz (1956) have shown this categorization to occur for the Negro stereotype under their experimental conditions. In their study, 10 Negro and 5 white photographs presented in a random order were rated by 58 judges on 15 attributes comprising the Negro stereotype. Negro photographs had been previously selected to range from Negroid to Caucasoid in appearance. The occurrence of categorization was confirmed by the fact that the presence of Caucasoid features did not reduce stereotyping.

Several limitations of this experiment are apparent.

1. The presence of the white photographs in the series, randomly interspersed with Negro ones, may have led judges to emphasize and contrast the supposed group attributes of whites with those of Negroes, thus accounting for categorization.

2. Categorization may have been a function of unfavorable response to Negro photographs rather than of stereotyping. Virtually all attributes used to represent the Negro stereotype were either unfavorable or, at best, of ambiguous affect. This condition does not permit drawing a distinction between unfavorable response and stereotyping.

3. Even though some of the Negroes had markedly Caucasoid features, all were spontaneously identified as Negroes. By extending the range of Negro-Caucasoid characteristics, the relationship between identification and categorization may be more fully explored, in a manner to be shown later.

The following are the major propositions tested here.

1. Identification of a photograph as Negro is sufficient to evoke categorical response; i.e., the judgment that the stimulus person possesses in full degree the attributes belonging to that category. Identification may occur spontaneously from the perception of Negroid physiognomy, or may be induced by experimental designation.

2. Categorization is a function of stereotyping rather than favorable-unfavorable reaction to the photographs.

3. Categorization occurs for judges with all degrees of prejudice; however, stereotyping is greater for the more prejudiced judges.

4. Stereotyped responses to Negro photographs are maximized when a frame of reference involving Negro-white comparisons is induced in a perceiver.

METHOD

Photographs. Issues for the years 1954–1957 of a Negro magazine with good reproduction were searched for appropriate photographs. Criteria included a full front view of the head, an absence of consensual facial expressions (e.g., a smile, laughter, a scowl), an absence of hats, pipes, cigarettes, glasses, etc. An attempt was made to obtain 15 photographs which were spaced at approximately equal intervals with respect to the Negroid-Caucasoid characteristics of their face and head. The 15 selected photographs were presented individually to 20 judges who were asked to rank them in the order of Negroidness. Utilizing median ranks as scale values according to the procedure suggested by Guilford (1936, p. 255), five photographs were deleted, leaving 10 with scale values at roughly equal intervals. These values, ranging from most Negroid to most Caucasoid, were: 1.3, 3.0, 4.5, 6.0, 7.5, 9.5, 10.5, 12.5, 13.5, 15.0. Five white photographs were also selected for use in establishing the various judgmental conditions.

Trait ratings. Attributes were chosen in four categories: favorable stereotype, unfavorable stereotype, favorable traits irrelevant to the stereotype, and irrelevant unfavorable traits. The main source from which these attributes were chosen was the repetition by Gilbert (1951) of the Katz and Braly study (1935). Since the favorable-unfavorable connotations of some of these previously employed traits were not clear, some of them were modified to produce a more definite connotation. For example, "happy-go-lucky" has a favorable connotation in terms of pleasantness and good humor, but also has an aura of irresponsibility

[1] This investigation was supported in part by a research grant (M-1892) from the National Institute of Mental Health, Public Health Service.

about it. Therefore, the term "jolly" was substituted. The 20 traits and their classification were as follows: *unfavorable stereotype:* lazy, dishonest, stupid, superstitious. *Favorable stereotype:* cheerful, enjoys music, jolly, deeply religious. *Favorable irrelevant:* sportsmanlike, patient, sincere, generous, loyal, courteous. *Unfavorable irrelevant:* stubborn, moody, quarrelsome, conceited, boastful, demanding.

Traits were rated on a seven-point scale, and from these ratings two basic scores were obtained for each photograph. One of these was the stereotype score which is the sum for all judges of the ratings on the four favorable stereotype attributes and the four unfavorable stereotype traits. The other was a favorableness score and was computed for each picture by subtracting the unfavorable irrelevant trait ratings from the favorable irrelevant ratings and summing across judges. These scores were transformed linearly so that they might be reported in terms of the seven-point rating scale.

Experimental conditions. In order to explore the comparative frame of reference hypothesis and the relationship between identification and categorization, the following four groups of judges were employed, each participating in a separate session.

Group	*N*	Picture Series
I	18	Unidentified Negro series
II	21	Unidentified Negro-white series
III	18	Identified Negro series
IV	15	Identified Negro-white series

For Groups I and II, photographs were presented without any comments as to race, Group I being exposed to the 10 Negro photographs, and Group II to the same photographs with five white ones randomly interspersed. For Group III, however, the experimenter announced that all of the 10 photographs to be projected would be of persons belonging to the Negro race, and that even though some of them may not look like Negroes, every one of them considers himself a Negro and has Negro blood. For Group IV, the experimenter designated each photograph as Negro or white as it was presented, and made the same remarks about the appearance of the photographs as he did for Group III.

In order to check upon the recognition of the racial identity of the person pictured in the photograph, at the end of the session photographs were briefly shown again and judges asked to indicate on their rating scales whether, when making their judgments, they had thought of the pictured individual as Negro or white. They were also permitted to use a question mark if they were not certain.

Finally, a 23-item Likert-type attitude scale was administered which yielded two scores: verbal stereotype and verbal prejudice. Two stereotype statements were provided for each of seven stereotype traits employed for picture ratings; e.g., "According to our standards, most Negroes are *dishonest*." The 9-item prejudice scale was composed largely of statements involving endorsement of segregation or group inferiority of the Negro.

RESULTS

For Groups I and II, exposed to unidentified photographs, only the seven most Negroid pictures were clearly thought to be Negro, with 38 of the 39 judges identifying them as such. The three most Caucasoid Negro photographs, however, were identified as those of persons belonging to the *white* race by 33 or more of the 39 judges. In the following discussion, therefore, a distinction must be made between the seven photographs recognized as Negroid and the three mistakenly thought to be white.

Categorization. The solid curve in Fig. 1 represents the mean stereotype rating for each photograph with judgmental conditions disregarded. It is apparent from an examination of the first seven points in these curves that, as long as the individual is recognized as a Negro, stereotyping does not decrease as the pictures become increasingly Caucasoid. In fact, the least stereotyped of the seven is the most Negroid photograph in the series. With two exceptions, moreover, the three photographs not recognized as Negro and the five white photographs have stereotype traits attributed to them to a much smaller extent than do the seven recognized Negro photographs. The difference between the mean stereotype score for the 10 Negro photographs and the five whites is significant ($F = 6.80$; $df = 1/13$; $P < .05$).

Apparently, as long as a Negro photograph is racially identified *by the perceiver*, the pictured individual is stereotyped in greater degree than a white person. Caucasoid appearance in itself does not lessen the extent to which stereotyped traits are attributed.

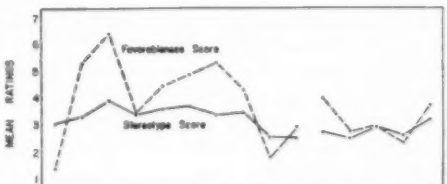

FIG. 1. STEREOTYPE AND FAVORABLENESS SCORES FOR EACH OF THE 15 PHOTOGRAPHS

Also noteworthy is that identification of the individual as Negro is not the only source of variance in the ratings; variance in stereotype scores from picture to picture is significant ($P < .001$). Individual differences in photographs are more dramatically represented by the favorableness scores (see Fig. 1). A consideration of the dotted-line curve leads to the conclusion that stereotyping cannot be identical with unfavorable reaction. Many of the Negro photographs are *more* favorably reacted to than the white ones; unlike stereotype scores, the mean score for Negroes is *higher* than for whites. Stereotype scores and favorableness scores do, however, have a high moderate correlation: Kendall's tau (Siegel, 1956) is .69 with $N = 10$ photographs and $P < .002$. This association is largely a reflection of the strong consensual personality impressions formed for each individual photograph. Some additional evidence demonstrating that stereotype and favorableness scores are *not* identical except for errors of measurement will be presented in a later section.

Identification and categorization. Although from the preceding discussion it would appear that racial identification is sufficient to produce categorization, there are several types of identification to be considered. The first seven photographs were spontaneously identified by judges in Groups I and II as Negro; presumably they had Negro characteristics in sufficient degree to enable such a decision. A few judges, however, did identify the three most Caucasoid photographs as Negroid rather than white; they may be expected to stereotype the three pictures to a greater degree than those judges assuming these photographs to be white. Table 1 presents the means and significance tests for the differences between the means for these judges. Those judges identifying the Caucasoid Negro photographs as Negro do indeed stereotype the two most Caucasoid of the three photographs to a significantly greater degree. The degree of stereotyping, moreover, is as great as that for Photograph I, which is the most Negroid in the series although not the most stereotyped. In part, perhaps, the reason that a significant difference is not obtained for Photograph J is that, even when identified as white, this picture has a fairly high stereotype score. One must ask whether the 10 judges identifying J, C, and D as Negro are highly prejudiced; their prejudice might account for their high stereotype scores. That this is not the case, however, is demonstrated by the fact that five of these judges are below the group mean on the prejudice scale, and five are above.

The effect of another kind of identification on stereotyping may also be determined. For Groups I and II, the three most Caucasoid photographs had been identified as white by almost all members. For Groups III and IV, however, all 10 photographs were identified as *Negro* by the experimenter. Of considerable

TABLE 1

STEREOTYPING OF NEGRO PHOTOGRAPHS AS A FUNCTION OF SPONTANEOUS RACIAL IDENTIFICATION BY THE PERCEIVER

Identification	Photograph					
	J		C		D	
	N	*M*	*N*	*M*	*N*	*M*
Negro	4	3.75	4	3.63	2	3.63
White	33	3.55	33	2.52	35	2.70
Mean difference		+0.20		+1.11		+0.93
P*		.2877		.0059		.0351

* Determined by a one-tailed application of the Mann-Whitney *U* test (Siegel, 1956).

TABLE 2

STEREOTYPING OF NEGRO PHOTOGRAPHS AS A FUNCTION OF RACIAL IDENTIFICATION BY THE EXPERIMENTER

Identification	Photograph		
	J	C	D
Judges High in Prejudice			
Unidentified (Groups I & II) ($N = 20$)	3.53	2.68	2.79
Identified (Groups III & IV) ($N = 17$)	3.90	3.02	3.22
Mean difference	+0.37	+0.34	+0.43
t	1.86	1.67	1.55
P*	.03	.05	.06
Judges Low in Prejudice			
Unidentified (Groups I & II) ($N = 19$)	3.63	2.66	2.70
Identified (Groups III & IV) ($N = 16$)	3.39	2.56	2.50
Mean difference	−0.24	−0.10	−0.20
t	1.10	0.42	0.91
P	.14	.34	.68

* Determined by a one-tailed t test.

interest is the response of these two sets of groups to the three photographs. Obviously, Groups III and IV should stereotype them to a greater extent.

If the means of Pictures J, C, and D, for Groups I and II are compared with the means for Groups III and IV, no significant difference in stereotyping is found. A reasonable assumption, however, is that only prejudiced judges should stereotype a Caucasoid photograph of a Negro more heavily when it is identified as to race by the experimenter. From Table 2, which presents the means for the judges high and low in prejudice, it is apparent that judges who are high in prejudice do indeed stereotype in greater degree when the three photographs are identified by the experimenter. That this is a result of stereotyping rather than a more unfavorable over-all reaction to the photographs when they are designated Negro is shown by the comparison of means for favorableness scores, as follows: High Prejudice judges in Groups I and II: J, 3.30; C, 1.65; D, 2.75; and those in Groups III and IV: J, 4.40; C, 1.05; D, 2.70. Obviously, there are no consistent differences in these means.

Prejudice and stereotyping. Already demonstrated is the point that prejudiced judges respond to the more Caucasoid photographs by stereotyping in greater degree when the experimenter identifies race, whereas nonprejudiced judges do not exhibit increased stereotyping under these circumstances. Other differences between these two groups of judges are apparent from Fig. 2. Judges high in prejudice stereotype *all* Negro photographs more heavily. A heterogeneous interaction effect is also apparent; greater differences between high and low prejudice judges occur for the seven most Negroid photographs than for the three most Caucasoid Negro photographs. An analysis of variance indicated that variance due to photographs, to prejudice, and to the interaction source (prejudice by photographs) were significant at the respective levels: .001, .001, and .01.

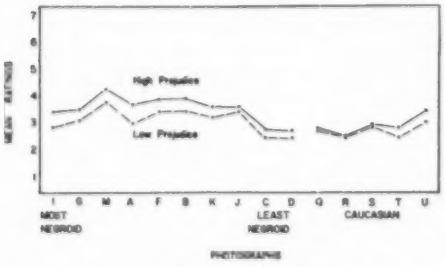

FIG. 2. STEREOTYPE SCORES BY PHOTOGRAPH FOR HIGH AND LOW PREJUDICE JUDGES

A graph for favorableness scores comparable to Fig. 2 is not presented, in the interests of conserving space. But an analysis indicated little difference between high and low prejudice groups. Variance for favorableness scores for photographs is, of course, significant ($P < .001$), but variances due to prejudice and to the interaction term prejudice by photographs are not significant ($P > .05$). These findings again suggest a demarcation between stereotyping and unfavorable reaction.

Also of interest are the comparisons among the two Likert scale scores, verbal stereotyping and verbal prejudice, and picture stereotyping. Verbal and picture stereotyping correlate .60, verbal stereotyping and prejudice also correlate .60, and prejudice and picture stereotyping correlate .36. All of these r's are significant at the .01 level with $N = 72$. From these correlations it is apparent that verbal and picture stereotyping have much in common, but are probably not identical, and likewise for the two verbal scales, prejudice and stereotyping. On the other hand, from the r of .36 it may be suggested that the tendency of an individual judge to stereotype photographs of Negroes is apparently determined by other factors in addition to his degree of prejudice toward Negroes. These conclusions must be regarded as tentative since the various scales have been constructed solely for this investigation, and have not been refined by conventional procedures of item analysis.

Frame of reference and categorization. The suggestion was made that presentation of a Negro-white series of photographs would lead to a greater degree of stereotyping than would a Negro series presented alone on the rationale that a comparative frame of reference would be induced by the Negro-white series of photographs. According to this hypothesis, stereotyping would occur in greater degree for Groups II and IV than for Groups I and III. From Table 3, however, it is apparent that although the predicted difference holds for

TABLE 3
VARIATION IN STEREOTYPE SCORES FOR NEGRO PHOTOGRAPHS AS A RESULT OF JUDGMENTAL CONDITIONS

	Judge Group					
	I	II	Diff.	III	IV	Diff.
High Prejudice Judges	3.73	3.59	−0.14	3.47	3.85	+0.38
Low Prejudice Judges	3.07	3.41	+0.34	3.20	3.31	+0.11

the judges low in prejudice, for the judges high in prejudice one of the comparisons is in the reverse direction. These contrasting results for high and low prejudice judges are represented by the interaction term for conditions by prejudice which is significant at the .001 level. It must be concluded that the predicted effect does not occur in a consistent manner, and that frame of reference has little to do with the main results reported to this point. Speculation as to the meaning of this interaction suggests a variety of alternative interpretations which, however, are fruitless to present, since the data do not favor a choice among them.

DISCUSSION

This study indicates that identification of a photographed person as Negro is sufficient to evoke the judgment that the individual possesses all the attributes commonly attributed to Negroes. Within certain limits, this phenomenon of *categorization* appears to be independent of the Negroid-Caucasoid facial attributes present in the photograph, and is also independent of favorable-unfavorable reaction to the photograph and the degree of prejudice characteristic of the judge. Even though less prejudiced judges stereotype all Negro photographs to a lesser degree than do more prejudiced judges, there is no tendency in either group to reduce stereotyping when photographs are more Caucasoid in appearance. These conclusions hold when the photograph is spontaneously identified as Negro. When, however, the Negro photograph is so Caucasoid in appearance that judges tend to identify it as white, those few judges who still identify the photograph as Negro continue to categorize. But designation of these highly Caucasoid photographs as Negro by the experimenter produces categorization only for judges above the median of a verbal prejudice scale, and not for those below. Apparently, the mere *knowledge* that a person is Negro is not in itself sufficient to produce stereotyping. Finally, the reasonable hypothesis that a mixed series of Negro and white photographs should produce a tendency to contrast the two groups in making stereotype ratings is not confirmed.

Whether or not categorization of Negroes occurs in interpersonal situations as well as in judgment of photographs depends upon a variety of additional factors not considered in the present study. Obviously, the definition of the situation and the relationship between perceiver and perceived would be important determining factors. On the other hand, there is a reasonable probability that categorization of other persons is important not only in the case of Negroes, but instead is a widespread process in person perception. Certainly it is not feasible to observe and grasp each individual in all his uniqueness and still carry on the smooth, fast-moving interaction with other persons which is characteristic of most everyday relationships. A reasonable assumption is that each perceiver has a set of person-categories (of which stereotypes are the most obvious) into which he classifies those with whom he interacts. The ordering of his behavior toward them is thereby simplified, in that the diversity of behavior which would be required if each person were unique is reduced to a finite number of categories. Further empirical work is indicated to explore categorization on a broader scale.

SUMMARY AND CONCLUSIONS

By presenting under different conditions to a variety of judge groups a series of 10 Negro photographs scaled to range widely from markedly Negro to markedly Caucasian, it was demonstrated that identification of a photograph as that of a Negro appears to be sufficient to evoke the judgment that the individual possesses all the attributes commonly assigned to Negroes. This phenomenon of categorization was found to be independent of Negroid-Caucasoid features, of favorable-unfavorable reaction to the photograph, of whether the picture series was Negro-white or Negro only, and of the degree of prejudice of the judge. Categorization breaks down only

when the photographs are so Caucasoid that they cease to be spontaneously recognized as Negro. Knowledge of racial identity provided by the experimenter for those pictures not otherwise recognized as Negro succeeds in producing categorization only for judges high in prejudice. This process of categorization, it was suggested, may be of general importance in person perception.

REFERENCES

GILBERT, G. M. Stereotype persistence and change among college students. *J. abnorm. soc. Psychol.*, 1951, **46**, 245–254.

GUILFORD, J. P. *Psychometric methods*. New York: McGraw-Hill, 1936.

KATZ, D., & BRALY, K. W. Racial prejudice and racial stereotypes. *J. abnorm. soc. Psychol.*, 1935, **30**, 175–193.

SECORD, P. F. Facial features and inference processes in interpersonal perception. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior*. Stanford: Stanford Univer. Press, 1958. Pp. 300–315.

SECORD, P. F., BEVAN, W., & KATZ, BRENDA. Perceptual accentuation and the Negro stereotype. *J. abnorm. soc. Psychol.*, 1956, **53**, 78–83.

SIEGEL, S. *Nonparametric statistics*. New York: McGraw-Hill, 1956.

(Received July 21, 1958)

JUDGMENT OF ADJUSTMENT BY PSYCHOLOGISTS, PSYCHIATRIC SOCIAL WORKERS, AND COLLEGE STUDENTS, AND ITS RELATIONSHIP TO SOCIAL DESIRABILITY[1]

MORTON WIENER
Clark University

ALBERT BLUMBERG
St. Vincent's Hospital, New York City

SARAH SEGMAN
Children's Hospital, Washington, D. C.

AND ALLAN COOPER
University of Rochester

IN THE clinical setting, it is often assumed that clinicians all have a similar, if not identical, set of constructs in terms of which they assess adjustment, and that they would agree in their denotative judgments about whether a set of behavior was or was not typical of an "adjusted" individual. Unless there is such agreement among clinicians, there is likely to be considerable between-judge unreliability whenever the clinicians must judge from observed behaviors or from self-referrent reports whether a patient is adjusted.[2]

Clinical experience suggests, however, that clinicians' implicit concepts of "adjustment" are often very vague and undefined, and that different clinicians or clinicians from different professional groups (e.g., psychologists, psychiatric social workers, psychiatrists) appear to emphasize different behaviors or self reports as characteristic of the adjusted individual. Thus, the reliability of judgments among clinicians about an individual's adjustment often seems considerably less than perfect.

A basic problem, then, is to try to determine the underlying model or the implicit construct system of adjustment which clinicians have. Further, it seems important to determine whether there is basic agreement among clinicians about the model. One way to proceed is to explore the clinician's judgments and expectations of what a well-adjusted individual reports about certain aspects of his behavior. The model can then be inferred from these judgments.

If this method of investigation proves feasible, it may help to make explicit the specific areas of agreement and of disagreement among clinicians in general, among clinicians from different professional groups, or among clinicians having different theoretical preferences.[3]

Judgments of clinicians may be obtained and examined in several ways. Rogers and Dymond (1954), for example, had two clinicians judge each item as appropriately positive or negative for adjustment. Inclusion in the scale was based upon the reliability of classification by the two judges and agreement by five other judges. This method resulted in a

[1] All of the data except the Social Desirability data were collected while Morton Wiener, Albert Blumberg, and Sarah Segman were on the staff at Central State Hospital, Indiana. The Social Desirability data were collected by Morton Wiener and Allan Cooper, while both were at the University of Rochester. The authors wish to thank C. L. Williams, Superintendent of Central State Hospital, and G. Richard Wendt, Department of Psychology, University of Rochester, for their support and encouragement. We express our thanks to Irving Weinstein, Earl Furlow, and Raymond Clausman for their helpful suggestions in the early planning phases of this experiment. We also wish to thank the many psychologists, social workers, and college students who participated as subjects and contributed their time and interest.

[2] It is clearly recognized that the problem of phenotypical versus genotypical behaviors must be taken into account. However, given a fixed set of behaviors or self-referrent statements, it should be possible for judges to agree about how a well-adjusted individual should *typically* respond.

[3] The original design of this study included plans for a programmatic investigation of several professional groups (e.g., psychiatric social workers, psychiatrists, teachers, nurses, judges, among others) and their judgment of the expected behavior an adjusted individual would report. These groups had been selected because their members are often required to make judgments about the adjustment of individuals. However, the results obtained from the psychologist and the social worker populations have led us to postpone further investigations. Data collected from psychiatric social workers and college students will therefore be presented as part of this study. If high agreement among judges can be demonstrated, and the study appears to be psychologically meaningful, the scale could be validated by comparing the scale score values of "adjusted" individuals with those of various "maladjusted" groups (e.g., neurotic, psychotic, and psychopathic).

scale of 74 items out of the 100 initially judged. Another method, which may increase the sensitivity of the scale, is to weight the relative importance of individual items. A Q-sort array appears to be ideal for accomplishing this weighting. The items in the array must be relevant to the concept of "adjustment," reflecting the population of possible items clinicians must judge in their everyday functioning. The 100 items used by Rogers and Dymond appear to meet these criteria.[4]

The purposes of this study, then, are (*a*) to determine whether a group of clinical psychologists agree about the relative importance of reported behaviors being characteristic of the "adjusted" individual; if agreement is found, (*b*) to examine the Adjustment scale derived from the combined judgments of the clinicians, (*c*) to compare this Adjustment scale to that derived by Rogers and Dymond, (*d*) to compare the agreement found among psychologists with the judgments of another professional group (psychiatric social workers) and with untrained college students, (*e*) to try to infer the implicit model of "adjustment" from the item sorts, and (*f*) to assess the relationship between the Adjustment scale value and the social desirability value of the items.

METHOD

Subjects. The *S*s were 44 clinical psychologists, 14 psychiatric social workers, and 21 college students. Twenty-eight of the 44 clinical psychologists and all the social workers and college students completed judgments for the Adjustment scale. The remaining 16 clinical psychologists judged the social desirability of each item.

Procedure. Each judge was given a set of the Rogers and Dymond items and completed the task independently and at his own convenience. The directions for the Adjustment scale judgments were as follows:

> We are interested in learning how you think a well-adjusted person would sort these cards. Please do this work by yourself. We are only interested in how you sort the cards.
>
> There are 100 cards in this deck. Each card has a statement of how people think, feel, act, etc. For any person they are more or less true, or not true, or in between. What you are to do is to sort these cards as you think a well-adjusted person would sort them. You are to sort the cards of the deck into 9 columns. To help you in this you have a small pack of 9 cards labeled "Column 1," "Column 2," etc. First put these 9 cards out on a table:
>
> Column 1 Column 2 Column 3···Column 8 Column 9
>
> Next, you are told on these column cards how many to put in each column and what each column stands for. For example, in Column 9, which is "most unlike the well-adjusted person," you are to place 4 cards. It is best not to cover one card with another, but to place them in columns below the column cards. In this way you can see all the cards, to read and compare and shift them, until you have placed them as you want to, with the right number in each column.
>
> As you go through the deck the first time you may put any number of cards in any column to start with, or you may put some cards to the left, some to the right, and some in the middle: then shifting them about as you read and compare them. Remember you are sorting them as you think a well-adjusted person would sort them.
>
> After you have completely finished the sorting, take the record sheet and write your name, age, sex, and occupation in the appropriate places. Each card you have sorted has a letter and a number (e.g., R1, or R 93, etc.). On the record sheet next to Column 9 are 4 blanks. Fill in the numbers of the 4 cards you have placed in Column 9 in the corresponding 4 blanks on the record sheet. Do the same for each of the other columns. Every blank should be filled in when you are finished recording. This is a good check to see if you have the right number of cards in each column.

The directions for the Social Desirability Q sort of the items were as follows:

> One of the things that has been found to influence subjects' answers to personality inventories is the social desirability or social acceptability of the items in the test. That is, people tend to answer items in a way that is considered the socially desirable way to be.
>
> We are actually trying to establish whether or not a group of items from a Q sort is socially desirable. We would like you to serve as a judge of the social desirability of this group of items.
>
> Please read each statement and decide whether you think *people in general* would consider that trait or behavior socially desirable or socially undesirable. Then decide, according to the distribution noted on the pink cards, where in the distribution of social desirability the item belongs.
>
> Remember you are to try to judge how *most people* or *people in general* would accept these traits or behaviors as socially desirable or undesirable.

RESULTS

The mean score and the variance of each of the 100 items of the Adjustment scale and of the Social Desirability scale were computed.[5]

[4] The authors wish to thank Carl R. Rogers for making the items available and for his permission to use them in this study.

[5] The means and variances for each of the 100 items for the three subject populations and the social de-

TABLE 1
CORRELATIONS AMONG THE THREE GROUPS FOR THE MEAN ADJUSTMENT SCALE VALUES OF THE 100 Q SORT ITEMS

	Social Workers	College Students
Psychologist	.94	.95
Social workers		.91

From these scores, the correlation coefficient of intragroup agreement for each of the three groups of judges was determined.[6] The intragroup agreement among the psychologists was .75; among social workers, .49; and among college students, .46. Examination of the variances for the items shows that both the social workers and the college students were more variable (and thus less reliable) in their judgments than the psychologists. The range of the score values of the 100 items is somewhat greater for the psychologists than for the other two groups: 1.50 to 8.39 for psychologists as compared with 2.43 to 7.93 for the psychiatric social workers and 2.43 to 7.71 for the college students.

Despite these differences in homogeneity of judgments, as examination of Table 1 indicates, the correlations between the mean values of the items by *S*s in the three populations were very high, falling between .91 and .95. These high correlations indicate that any disagreements among the judging populations about the relative importance of an item for adjustment are minor. Disagreements constitute at most a slight variation in the relative importance attributed to specific items. However, if one assumes that the 10 highest and the 10 lowest items are reliably ranked, some subtle differences among the groups are noted. Table 2, in which the 10 most characteristic items of the adjusted individual are listed, shows that, in general, the college students appear to emphasize a positive self-attitude and a social, other-directed adjustment. The social workers place emphasis upon rationality and initiative as well as self-acceptance and good interpersonal relationships. The psychologists stress responsibility, self-acceptance, and good interpersonal relationships. Table 3 shows the 10 items considered least characteristic of the well-adjusted person for the three subject groups. The psychologists and social workers agree completely with each other about the specific 10 items most unlike the well-adjusted individual. The college students agree only with six of the items and add four others. In essence, helplessness and a general feeling of disturbance and problems in the area of one's own sexuality, in addition to non-self-acceptance are important attributes for the professionals. Hostility expressions, feelings of inferiority, unreliability of behavior, and non-self-acceptance are the important items for the college students.

Rogers and Dymond report 37 items as positive value contributions to their Adjustment scale and 37 as negative value contributions. The Adjustment scale derived from the Q sorts of the 28 psychologists shows less than perfect agreement with the Rogers and Dymond classification of items. On the Q sort, a score of up to 4.43 includes the 37 most positive items, and a score above 5.64 includes the 37 most negative items. Comparison of the common items shows that only 33 of the 37 positive items and only 29 of the 37 negative items occur in both scales. Using a forced-normal Q sort, it would be possible under compete agreement among judges to have 40 items with a score of 4.0 or below and 40 items with a score of 6.0 or above. In the present study there were only 31 items in each of these categories. Only 25 positive and 26 negative items are common to this list and to Rogers and Dymond's 74 reliable items.

The mean scores and variances of the social desirability ratings for the 100 items derived from the judgments of the 16 psychologists

sirability value of the items have been deposited with the American Documentation Institute. Order Document No. 5966, from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

[6] The intercorrelation within each subject group was computed by the formula (Cronbach, 1951):

$$r_{pp} = \frac{1}{N-1} \cdot \frac{NVxi - Vx}{Vx}$$

N = number of subjects
Vxi = variance of the means of the items averaged over all the subjects; and Vx = the variance of the Q sort.

TABLE 2

THE TEN ITEMS CONSIDERED MOST CHARACTERISTIC OF THE WELL-ADJUSTED PERSON FOR THE THREE SUBJECT GROUPS, AND THE TEN ITEMS HIGHEST ON THE SOCIAL DESIRABILITY SCALE OF THE PSYCHOLOGISTS

Psychologists	Social Workers	College Students	Social Desirability
I am a responsible person 3.5	I am a responsible person (8)	I am a responsible person (3)	I am a responsible person (3)
I feel emotionally mature (3.5)	I feel emotionally mature (4.5)	I feel emotionally mature (8.5)	I feel emotionally mature (1)
I can usually live comfortably with people around me (2)	I can usually live comfortably with people around me (1.5)	I can usually live comfortably with people around me (3)	
I have a warm emotional relationship with others (1)	I have a warm emotional relationship with others (1.5)		I have a warm emotional relationship with others (2)
I can accept most social values (9)	I can accept most social values (6.5)		
I usually like people (5.5)	I usually like people (4.5)		
I take a positive attitude towards myself (8)	I take a positive attitude towards myself (9.5)		
I feel adequate (5.5)	I feel adequate (3)		
I am liked by most people who know me (10)		I am liked by most people who know me (8.5)	I am liked by most people who know me (4)
I am self-reliant (7)		I am self-reliant (1)	
	I have initiative (9.5)	I have initiative (3)	I have initiative (9)
	I am a rational person (6.5)		
		I understand myself (5)	
		I am a good mixer (10)	I am a good mixer (9)
		I am poised (6)	
		I am ambitious (7)	
			I have an attractive personality (5)
			I am likeable (6.5)
			I am a hard worker (6.5)
			I am intelligent (9)

Note.—The numbers in parentheses indicate the item's rank.

TABLE 3

THE TEN ITEMS CONSIDERED LEAST CHARACTERISTIC OF THE WELL-ADJUSTED PERSON FOR THE THREE SUBJECT GROUPS, AND THE TEN ITEMS LOWEST ON THE SOCIAL DESIRABILITY SCALE OF THE PSYCHOLOGISTS

Psychologists	Social Workers[a]	College Students	Social Desirability[a]
I want to give up trying to cope with the world (2.5)	I want to give up trying to cope with the world (6.5)	I want to give up trying to cope with the world (9.5)	I want to give up trying to cope with the world (7.5)
I am no one; nothing seems to be me (1)	I am no one; nothing seems to be me (1.5)	I am no one; nothing seems to be me (3.5)	I am no one; nothing seems to be me (6)
I despise myself (2.5)	I despise myself (3)	I despise myself (2)	I despise myself (1)
I am a failure (7)	I am a failure (5)	I am a failure (3.5)	I am a failure (2)
I feel hopeless (5)	I feel hopeless (1.5)	I feel hopeless (7.5)	I feel hopeless (10)
I am worthless (4)	I am worthless (4)	I am worthless (1)	I am worthless (3)
I have a feeling of hopelessness (9)	I have a feeling of hopelessness (10.5)		
I feel helpless (8)	I feel helpless (10.5)		
I really am disturbed (6)	I really am disturbed (8)		I really am disturbed (4)
I dislike my own sexuality (10)	I dislike my own sexuality (6.5)		I dislike my own sexuality (10)
	I am afraid of sex (9)		
		I am a hostile person (7.5)	
		I feel inferior (6)	
		I just don't respect myself (5)	I just don't respect myself (10)
		I am unreliable (9.5)	I am unreliable (7.5)
			I am much like the opposite sex (5)

Note.—The number in parenthesis indicates the item's rank.

[a] These two groups had two items tied for 10th ranking, thus have 11 items listed.

were computed.[7] The intragroup correlation coefficient determined from these scores is .62. The range of scores is from 1.87 to 8.44. The correlation of the Adjustment scale scores and the Social Desirability scale score is .88. In general, there is very high agreement in the placement of items for the Adjustment scale and for Social Desirability.

Examination of the 10 highest and 10 lowest ranked scores for the two scales (Tables 2 and 3) shows that there is agreement for 12 of the 20 items between the Psychologists' Adjustment scale and their social desirability rankings. These 10 highest and lowest ranked items on the Social Desirability scale are no more common to either of the Social Workers' Adjustment scale or the College Students' Adjustment scale than to the Psychologists' Adjustment scale.

DISCUSSION

The results of the Q-sort arrays by the psychologists indicate there is considerable agreement among experienced clinicians about the specific behaviors that an adjusted person is expected to report. However, the high correlation of the Adjustment scale judgments with the Social Desirability judgments possibly indicates that in the absence of knowledge about the underlying genotypical factors, judges resort to a concept of social acceptability of specific behaviors. It is noteworthy that no questions were raised by judges from any of our population groups about specific characteristics of the adjusted individual they were to characterize, e.g., sex, age, education, marital status, or other possible genotypical factors. Either the behaviors in our set of behaviors are nonspecific for those factors, often considered basic for evaluating adjustment, or our judges made their own idiosyncratic decisions about the kind of individual whose sorting they were to try to evaluate. If this latter hypothesis is correct, it is possible that higher interjudge agreement would have been found for the Adjustment scale judgments if the "adjusted" individual had been more fully specified. However, the interjudge agreement for the Adjustment scale was no lower than that for the judgments of the Social Desirability of the items (Adjustment scale .75, Social Desirability .62).

These findings suggest that psychologists may have a stereotyped concept about the kinds of behavior they expect an adjusted individual to report. The high correlation between the Adjustment scale scores and the Social Desirability scale scores, however, indicates that the Social acceptability of the behaviors seems to be a major determinant of the Q sort.

One can argue that a well-adjusted person does have behavior traits which are socially desirable, and that an adjusted individual's report that these are characteristic of himself underlies our results, not the clinician's tendency to equate social desirability with adjustment.

It is conceivable that those people who are well-adjusted are those who share the model of the well-adjusted person held by professional people (which also happens to be socially desirable behavior) and whose concept of "normality" is congruent with their own behavior. Conversely, maladjusted people are those who may or may not share the model of normality, but if they do, it is not congruent with their nonverbal behavior. Two specific issues are inherent in this hypothesis. It must be shown that (*a*) adjusted and matched nonadjusted populations sort differently, or (*b*) if they sort the same way, the behavior of the adjusted population is congruent, while the behavior of the maladjusted is noncongruent with their verbal reports. If Condition (*a*) is empirically demonstrated, then social desirability is not an artifact of an adjustment scale. However, if Condition (*b*) is the case, then some new procedure for assessing adjustment will need to be developed. If the latter interpretation is correct, then it is important for clinicians to establish evaluation procedures which control for social desirability if they wish to evaluate adjustment or changes in adjustment. Unless this is done, it could not be determined whether an *S*'s adjustment had improved or whether he was simply verbalizing in more socially acceptable ways but without congruence to his own nonverbal behaviors. The data in this study do not resolve this problem.

[7] See Footnote 5.

The Adjustment scale findings in this study indicate further that, although Rogers and Dymond's scale does to some extent represent expectations of report from an adjusted *S*, *S*s who score as "adjusted" are also reporting essentially socially acceptable behaviors, and their scores can therefore be highly influenced by a set to represent themselves as socially acceptable. Moreover, the data suggest that the Adjustment scale derived by Rogers and Dymond and the scale derived by a forced normal distribution of items result in different "Adjustment" scales.

It may be argued that a forced Q sort in which items must be ordered in a positive and negative array should result in greater variability of scale values than if the items are put in only two categories, plus or minus. However, since Rogers and Dymond had 26 items which were judged as neither positive nor negative for adjustment, we should expect that these unreliable items would fall in our middle (unreliable) categories and that the Rogers and Dymond relevant items would fall at the extremes of the forced-choice scale. This, however, was not the case. Comparison of the two scales indicates that only 62 of the 100 items (31 high and 31 low) can be considered as reliably judged as positive and negative behavior attributes, contrasting with the 74 suggested by Rogers and Dymond. In addition, of the 62 items, only 25 positive and 25 negative are common to the adjustment scale in this study and to Rogers and Dymond's Adjustment scale. The remaining 12 items which were found reliable here are items which Rogers and Dymond's judges did not consider particularly relevant for adjustment. However, because of the limited number of cards any judge may place in any column, there is some question whether a forced Q-sort array would not necessarily decrease the number of items with scale values below 4.0 and above 6.0 which can be considered reliable. In any case, the failure to find greater consistency of item placement with the two methods of judgment raises some question about the need to investigate further the problem of establishing an Adjustment scale by either method.

The apparent greater variability of the other populations in this study, compared to the psychologists, is difficult to explain. It is not unexpected that a nonprofessional group, college students, would show greater variability of judgment about the relative importance of particular behaviors for adjustment, although the mean score values of their judgments for items are highly similar to those of the other groups. However, our data do not indicate any single item or group of items which may have resulted in a systematic increase in variability of judgment by the social worker group.

The findings of Kogan, Quinn, Ax, and Ripley (1957) and of Edwards (1953) and Edwards and Horst (1953), on social desirability and adjustment scores are highly consistent with the results of this study. Because of the high similarity between the Adjustment scale scores and the Social Desirability of the items, it is necessary to investigate carefully the validity of the derived Adjustment scale to find out the extent to which it measures adjustment independent of reports of social desirability. Does it indicate the behavior an adjusted individual is likely to *report* in contrast to that reported by a maladjusted individual, or does it merely represent the behaviors a defensive *S* is likely to report? The findings in this study suggest that clinicians need to be quite careful in attempting to apply scales of adjustment (e.g., Rogers & Dymond, 1954 or Buss & Guerjoy, 1957) without trying to control or assess the overlap of these scales with the dimension of social acceptability. Cowen (1958), for example, reports a correlation of .91 between his *S*s' judgment of social desirability and the "normal" scale value of a large number of trait items from the Buss and Guerjoy list.

SUMMARY

An attempt was made (*a*) to determine whether a group of clinical psychologists can agree about the relative importance of reported behaviors being characteristic or not characteristic of the "adjusted" individual, (*b*) to examine the Adjustment scale derived from the combined judgments of the clinicians, (*c*) to compare the scale to that derived by Rogers and Dymond, (*d*) to compare the agreement found among psychologists with the judgments of another professional group (psychiatric social worker) and untrained

college students, and (*e*) to investigate the concept of "normal," inferred from the denotative judgments of these items, and (*f*) to assess the relationship between the Adjustment scale value and the social desirability value of the items.

A Q sort of the 100 items of Rogers and Dymond was completed by 28 psychologists, 14 psychiatric social workers, and 21 college students to represent how a "well-adjusted" individual would sort the items. Sixteen additional psychologists also sorted the items for Social Desirability. The findings were that psychologists showed considerable agreement: .75 for the Q-sort array for Adjustment and .62 for the array for Social Desirability of the items. The correlation of .88 between the mean scale scores for the two arrays implies little independent variance for the two concepts. Although there was more variability of judgments among the social workers and among the college students, the mean Adjustment scale score values for the 100 items were highly correlated between population groups (correlations from + .91 to +.95). The implications are discussed, and care is suggested in any attempt to develop or use Adjustment scales or concepts of adjustment without assessing the overlapping contribution of the Social Desirability dimension until such time as empirical evidence is available that adjusted and nonadjusted populations respond differently to Adjustment scales despite this overlap of Social Desirability items; or that socially desirable behavior is found to be congruent with adjusted population reports and noncongruent with that of a matched nonadjusted population.

REFERENCES

BUSS, A. H., & GUERJOY, H. The scaling of terms used to describe personality. *J. consult. Psychol.*, 1957, **21**, 361–369.

COWEN, E. L. Sex differences in Social Desirability ratings of trait description terms. Paper read at Eastern Psychological Association, Philadelphia, April 1958.

CRONBACH, L. J. Correlations between persons as a research tool. Unpublished paper read at the Symposium, Research Methods in Clinical Psychology at Washington University, St. Louis, Missouri, March 10, 1951. Pp. 1–17. (Mimeo.)

EDWARDS, A. L. The relationship between the judged desirability of a trait and the probability that the trait will be endorsed. *J. appl. Psychol.*, 1953, **37**, 90–93.

EDWARDS, A. L., & HORST, P. Social desirability as a variable in Q-technique studies. *Educ. psychol. Measmt.*, 1953, **13**, 620–625.

KOGAN, W. S., QUINN, R., AX, A. J., & RIPLEY, H. S. Some methodological problems in the quantification of clinical assessment by Q array. *J. consult. Psychol.*, 1957, **21**, 57–62.

ROGERS, C. R., & DYMOND, ROSALIND F. *Psychotherapy and personality change.* Chicago: Univer. Chicago Press, 1954.

(Received July 23, 1958)

SOME PERSONALITY DETERMINANTS OF THE EFFECTS OF PARTICIPATION[1]

VICTOR H. VROOM

Survey Research Center, University of Michigan

PSYCHOLOGISTS have long realized the importance of both environmental and personality variables in the explanation of behavior. Theorists have employed a variety of terms to describe the necessity of using both sets of concepts. Lewin (1951), for example, illustrates this dual focus in his statement that "behavior (B) is a function (F) of the person (P) and of his environment (E), B = F (P, E)" (p. 239).

There has, however, been a tendency for investigators in social psychology to concentrate on one or the other of these sets of variables in their explanation of social phenomena. Some emphasize personality, conceived as the relatively enduring psychological properties of an individual, as the locus of the basic causes of behavior, while others look to environmental variables such as group structure, communication, and role. Few have investigated environmental and personality determinants of behavior simultaneously.

The implications of this point of view for problems of leadership have been described by a number of writers (Gibb, 1954). The general conclusion is that leadership cannot be regarded as a unitary trait and must be evaluated in terms of a number of other variables including the attitudes, needs, and expectations of the followers. The most effective behavior in dealing with individuals with certain personality characteristics may be ineffective in dealing with persons differently predisposed.

A similar point is made by those who argue for the adaptive nature of leadership. After reviewing research on the effectiveness of different methods of supervision in industry, Likert (1958) reaches the following conclusion:

> Supervision is, therefore, always an adaptive process. A leader, to be effective, must always adapt his behavior to fit the expectations, values, and interpersonal skills of those with whom he is interacting (p. 327).

The "authoritarian-democratic" continuum represents one aspect of leadership that has received much attention. In discussing studies dealing with this dimension Krech and Crutchfield (1948) suggest that its effects may vary from culture to culture:

> All the experimental evidence to be reported has been obtained by the study of so-called "authoritarian" and "democratic" leadership situations in our democratic culture. It is entirely possible that similar studies in other cultures might yield different results. The advantages for morale, the experiments find, seem to be with the democratically led group, but in an autocratic culture the reverse might possibly hold true (p. 423).

Despite frequent speculations that the superior effects of democratic leadership are specific to certain personality types or cultures, relatively little research has been done on this problem. The few studies which have been carried out (French, Israel, & Ås, in press; Sanford, 1950; Tannenbaum & Allport, 1956) have produced positive results. The task remains to determine the nature of the personality variables and their manner of interaction with democratic leadership.

The major purpose of the present study is to determine whether the effects of one aspect of the democratic leadership process—participation in decision-making—vary with the personality structure of the follower. Previous research has demonstrated that participation, or conceptually similar variables, has positive effects on the attitudes and performance of the participants. The general hypothesis of this study is that participation interacts with certain personality characteristics of the participant in determining both attitudes and performance.

The personality variables thought to be relevant in determining an individual's response to participation are (*a*) need for inde-

[1] This paper is based on a dissertation submitted to the Department of Psychology, University of Michigan, in partial fulfillment of the requirement for the degree of doctor of philosophy. The writer is indebted to J. Atkinson, J. R. P. French, Jr., and F. Mann for their helpful comments and criticisms, and to the United Parcel Service for their financial support of this research.

pendence and (*b*) authoritarianism. Participation is hypothesized to have more positive effect on the attitudes and performance of persons with strong than weak independence needs and to have less positive effect on authoritarians than equalitarians. The relevance of need for independence in determining a person's reactions to participation was suggested by McGregor (1944), while Sanford (1950) has shown that authoritarians say they prefer more strongly directive kinds of leadership.

Participation, the independent variable in this study, has been used in a number of ways and has seldom been clearly defined. The present investigation employs the definition put forth by French, Israel, and Ås (in press): a process of joint decision-making by two or more parties in which the decisions have future effects on those making them. The amount of participation by any individual is the amount of influence he has on the decisions and plans agreed upon.

It is important to distinguish between psychological participation, or the amount of influence he perceives he has on decision-making and objective participation, or the amount of influence he actually does have on decision-making. If perception is veridical, the amount of psychological participation equals the amount of objective participation. Frequently, however, they differ as a result of the influence of processes such as the effects of needs on perception. Concern here is limited to psychological participation and, unless otherwise noted, the term "participation" is used to designate this variable.

METHOD

This study was carried out in a large company whose basic function is the delivery of small parcels and packages from department and other retail stores to private residents. The *S*s were 108 first, second, and third line supervisors in the company's two largest plants.

Measures were obtained on each of the following variables:

1. Psychological participation. This index is derived by summing the responses of each supervisor to the following questions:

(*a*) In general, how much say or influence do you have on what goes on in your station? (*b*) Do you feel you can influence the decisions of your immediate superior regarding things about which you are concerned? (*c*) Does your immediate superior ask your opinion when a problem comes up which involves your work? (*d*) If you have a suggestion for improving the job or changing the setup in some way, how easy is it for you to get your ideas across to your immediate superior?

Each of these questions was answered by checking the most applicable alternative on a five-point scale. Scores ranging from 1, representing low participation, to 5, representing high participation, were assigned to each question and total scores obtained for each person by summing his scores for the four items.

The test-retest reliability of this index over a seven-month period is .61 for 91 supervisors. When 14 supervisors who changed either their position or their superior during this period were removed from this group, the reliability coefficient increased to .63. The correlation for the transferees is .44.

2. Attitude toward the job. The measure of attitude toward the job consists of the following items:

(*a*) How well do you like supervisory work? (*b*) How much of a chance does your job give you to do the things that you are best at? (*c*) How good is your immediate superior in dealing with people?

Each of the questions calls for the respondent to check the most appropriate answer on a five-point scale. Scores of 1 to 5 were assigned to each question and a total score obtained by adding over the three questions.

The test-retest reliability of this index over a seven-month period was computed for 91 supervisors and found to be .66. When 14 supervisors who had changed their superior or their position within the organization during this period were removed, the reliability coefficient was increased to .75. The reliability for the transferees was .06.

3. Need for independence. The measure of need for independence used in this study consists of 16 questionnaire items.[2] Some of these items refer to the frequency with which the *S* regularly engages in independent behavior (e.g., "How often do you find that you can carry out other people's suggestions without changing them any?"), while others deal with the satisfaction that he gets from this behavior (e.g., "When you have a problem, how much do you like to think it through yourself without help from others?"). The items are adapted from a larger number employed by Tannenbaum and Allport (1956). Each item required the *S* to check one of five alternatives.

No data were available concerning the reliability of this measure. However, the test-retest reliability over a seven-month period of a short form, made up of eight items selected on the basis of item analysis from the original items, was found to be .61 for 90 supervisors.

---

[2] Tables showing the items, scoring keys, means, and standard deviations for the need for independence, authoritarianism and job performance measures have been deposited with the American Documentation Institute. Order Document No. 5965 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

4. Authoritarianism. The degree of authoritarianism of the *Ss*[3] is measured by responses to 25 items from Forms 40 and 45 of the F scale developed by Adorno, Frenkel-Brunswick, Levinson, and Sanford (1950). In keeping with the rest of the questionnaire, *Ss* were asked to check their degree of agreement with each of the statements on a five-point scale, unlike the six-point scale usually used.

5. Job performance. Ratings of the job performance of 96 of the supervisors in the sample were completed by the immediate superior of the man being rated and reviewed by one other person who was acquainted with his work. These ratings consisted of two scores—over-all performance and summary appraisal. The over-all performance rating was obtained from a modification of the forced-choice merit rating technique. It consisted of ten sets of five statements, each describing some aspect of job behavior. The rater is asked to rank order the statements within each group in terms of the degree to which they describe accurately the man whose performance they are rating. The rankings on the ten sets of statements are scored to yield a single over-all performance score. The summary appraisal rating was of the graphic type. The rater is asked to check on a five-point scale his general evaluation of the degree to which the individual meets the demands of his job.

The hypothesis that participation has different effects on persons with different personality characteristics was tested in the following manner: The sample was divided into three approximately equal groups on the basis of their scores on each of the personality variables. Pearson product-moment correlation coefficients were then computed between participation and both attitude toward the job and job performance for the entire sample and for each of the subgroups.

TABLE 1

RELATIONSHIP BETWEEN PSYCHOLOGICAL PARTICIPATION AND ATTITUDE TOWARD THE JOB FOR PERSONS WITH DIFFERENT PERSONALITY CHARACTERISTICS

	Number of Cases	r
Total Group	108	.36***
1. High Need Independence	38	.55***
2. Moderate Need Independence	32	.31**
3. Low Need Independence	38	.13
diff (1, 3) t = 2.04[a]		
P = .02		
diff (1, 2) t = 1.20[a]		
P = .12		
diff (2, 3) t = —[a]		
P = —		
4. High Authoritarian	34	.03
5. Moderate Authoritarian	34	.35**
6. Low Authoritarian	39	.53***
diff (4, 6) t = 2.33[a]		
P = .01		
diff (4, 5) t = 1.36[a]		
P = .09		
diff (5, 6) t = —[a]		
P = —		

** $P < .05$.

*** $P < .01$.

[a] Indicates that the difference between correlations is in the predicted direction; t ratios over 1.00 are shown. Inasmuch as the direction of results has been specified in our hypotheses, one-tailed tests of significance have been performed.

RESULTS

The findings presented first deal with the effects of participation in decision-making on the attitudes toward the job of persons with different personality characteristics. More positive relationships between participation and attitude toward the job are predicted for persons who received relatively high scores on the need for independence measure and low scores on the authoritarianism measure than for persons at the opposite ends of these two scales.

The data in Table 1 support this prediction. The correlation between the measure of participation and attitude toward the job is significantly positive, confirming past findings. Significant differences are found, however, between the magnitude of the correlations for the different personality groups. As predicted, the most positive relationships between psychological participation and attitude toward the job are found for persons high in need for independence and low in authoritarianism. Both correlations are significant at the .01 level of confidence. The least positive relationships are found for persons low in need for independence and high in authoritarianism. Neither of these correlations is significantly different from zero. The differences between correlations for high and low groups on both personality variables are statistically significant.

The data are interpreted as meaning that the attitudes toward the job of low authoritarian persons and of persons with high independence needs are favorably affected by opportunities to participate in making decisions in their jobs. On the other hand, the attitudes of highly authoritarian individuals and of individuals with low independence needs are relatively unaffected by this experience.

The correspondence between the findings for need for independence and authoritarianism suggests the possibility that the measures of

[3] One *S* declined to complete the F scale, limiting the sample on this variable to 107 persons.

TABLE 2
RELATIONSHIP BETWEEN PSYCHOLOGICAL PARTICIPATION AND RATINGS OF JOB PERFORMANCE FOR TOTAL GROUP AND FOR PERSONS WITH DIFFERENT PERSONALITY CHARACTERISTICS

	N	Pearson *r*'s Between Participation and Supervisors' Rating on:	
		Over-all Performance	Summary Apraisal
Total Group	96	.20**	.20**
1. High Need Independence	33	.33**	.25*
2. Moderate Need Independence	28	.19	.33**
3. Low Need Independence	35	.06	−.01
diff (1, 3) *t* =		1.12[a]	1.08[a]
P =		.13	.14
diff (1, 2) *t* =		—[a]	—
P =		—	—
diff (2, 3) *t* =		—[a]	1.30[a]
P =		—	.10
4. High Authoritarian	30	−.08	−.06
5. Moderate Authoritarian	33	.28*	.23*
6. Low Authoritarian	32	.28*	.27*
diff (1, 3) *t* =		1.37[a]	1.26[a]
P =		.08	.10
diff (1, 2) *t* =		1.42[a]	1.12[a]
P =		.08	.13
diff (2, 3) *t* =		—	—[a]
P =		—	—

* $P < .10$.
** $P < .05$.
[a] In the predicted direction.

these two variables have a high negative relationship with one another. This possibility was tested by intercorrelating the two measures. The Pearson product-moment correlation coefficient between authoritarianism and need for independence is −.11 for 107 *Ss*. When age, occupational level, and education are partialled out from this relationship, the correlation is changed to .02, indicating that the two measures are independent.

Table 2 shows the intercorrelations between the measure of participation and the two measures of job performance for the entire sample and for the six subgroups. The significant correlations for the entire sample support previous findings that participating in making decisions in a job generally has positive effects on the job performance of the participant. In addition, some support is provided the hypothesis that the effects of participation on performance are a function of the need for independence and authoritarianism of the participant. Although none of the differences between correlations is significant, all of the high–low differences and most of the other differences are in the predicted directions.

The use of the correlational methods of field studies instead of the more precise techniques of laboratory experimentation increases the possibility that the results may be attributable to failure to control for relevant variables. The effects of the age, education, and occupational level of the *Ss* in the present sample were determined by partialling out these variables from the relationships between participation and the three dependent variables for the entire sample and for each of the six personality classifications. Table 3 shows the third-order partial correlation for each of these relationships.

A comparison of Table 3 with Tables 1 and 2 shows that partialling out the effects of background variables generally increases the magnitude of the differences between correlations and provides increased support for the hypothesis of interaction between participation and personality. All of the differences in correlations are in the predicted direction, and five of the six high–low differences are significant at the .05 level.

DISCUSSION

The present study corroborates previous findings that participation in decision-making has positive effects on attitudes and job performance. It further demonstrates that the magnitude of these effects is a function of certain personality characteristics of the participants. Authoritarians and persons with weak independence needs are apparently unaffected by the opportunity to participate in making decisions. On the other hand, equalitarians and those who have strong independence needs develop more positive attitudes toward their job and greater motivation for effective performance through participation.

There is no evidence of any unfavorable effects of participation either on attitudes or on performance. It should be noted, however,

TABLE 3

RELATIONSHIP BETWEEN PSYCHOLOGICAL PARTICIPATION AND ATTITUDE TOWARD THE JOB AND RATINGS OF JOB PERFORMANCE FOR PERSONS WITH DIFFERENT PERSONALITY CHARACTERISTICS WITH AGE, EDUCATION, AND OCCUPATIONAL LEVEL HELD CONSTANT

		Pearson r's Between Participation and:		
		Attitude toward the Job	Over-all Performance	Summary Appraisal
Total Group		.27*** (108)	.21** (96)	.20** (96)
1. High Need Independence		.51*** (38)	.51*** (33)	.42** (33)
2. Moderate Need Independence		.25* (32)	.18 (28)	.33** (28)
3. Low Need Independence		−.04 (38)	.04 (35)	.00 (35)
diff (1, 3)	t =	2.40[a]	1.93[a]	1.61
	P <	.01	.03	.05
diff (1, 2)	t =	1.15[a]	1.31[a]	—[a]
	P =	.12	.10	—
diff (2, 3)	t =	1.15[a]	—[a]	1.26[a]
	P =	.12	—	.10
4. High Authoritarian		.09 (34)	−.13 (30)	.14 (30)
5. Moderate Authoritarian		.35** (34)	.24* (33)	.18 (33)
6. Low Authoritarian		.50*** (39)	.33** (32)	.26* (32)
diff (4, 6)	t =	1.77[a]	1.68[a]	—[a]
	P =	.04	.05	—
diff (4, 5)	t =	1.04[a]	1.32[a]	—[a]
	P =	.15	.09	—
diff (5, 6)	t =	—[a]	—[a]	—[a]
	P =	—	—	—

* $P < .10$.
** $P < .05$.
*** $P < .01$.
[a] Indicates that the difference between correlations is in the predicted direction.

that the sample of supervisors used in this study is not representative of workers in general. It is possible that nonsupervisory employees might be more authoritarian and have weaker independence needs which might lead, in the extreme, to negative consequences of participation.

These results suggest the inadequacy of generalizations concerning the effects of participation. Studies that ignore the interaction of participation and personality yield relationships that are nothing more than average effects of participation for all the persons in the group. The statistic used to estimate the degree of relationship underestimates the effects of participation on some persons and overestimates the effects on others.

A word of caution should be injected here. The measure of participation used in this study was based on *S*s' reports and conforms with what we have defined as psychological participation. Enough is known about distorting influences in complex perceptions of this sort to make it impossible to infer that our measure of psychological participation corresponds with objective participation. Since no data are available on objective participation, extension of the present findings to cover the latter variable will require further research.

The results suggest that an adequate theoretical explanation of the effects of participation in decision-making should include a consideration of the influence of personality variables that interact with participation. The present study also gives general support to a situational theory of leadership and indicates the possible value in simultaneous examination of environmental and personality variables.

SUMMARY

The primary purpose of this study was to determine the effects of participation in decision-making on persons with different personality characteristics. It was hypothesized that equalitarians and individuals with strong independence needs would be more positively affected by the opportunity to participate in making decisions than authoritarians and persons with weaker independence needs.

The findings corroborated previous evidence that participation generally has positive effects on both attitudes and job performance. Hypotheses were also confirmed that the magnitude of these effects is a function of certain personality characteristics of the participant. Authoritarians and persons with weak independence needs are apparently unaffected by the opportunity to participate in making decisions. On the other hand, equalitarians and those who have strong independence needs develop more positive attitudes toward their jobs and increase in performance through participation.

This study suggests that an adequate theoretical explanation of the effects of participation in decision-making should include a consideration of the influence of personality variables. It also indicates the possible value of investigating the joint effects of leader and follower characteristics.

REFERENCES

ADORNO, T., FRENKEL-BRUNSWICK, ELSE, LEVINSON, D. J., & SANFORD, N. *The authoritarian personality.* New York: Harper, 1950.

FRENCH, J. R. P., JR., ISRAEL, J., & ÅS, D. An experiment on participation in a Norwegian factory. *Hum. Relat.*, in press.

GIBB, C. A. Leadership. In G. Lindzey (Ed.), *Handbook of social psychology.* Cambridge: Addison-Wesley, 1954. Pp. 877–920.

KRECH, D., & CRUTCHFIELD, R. S. *Theory and problems in social psychology.* New York: McGraw-Hill, 1948.

LEWIN, K. Behavior and development as a function of the total situation. In D. Cartwright (Ed.), *Field theory in social science.* New York: Harper, 1951. Pp. 238–303.

LIKERT, R. Effective supervision: An adaptive and relative process. *Personnel Psychol.*, 1958, **11:** 317–332.

MCGREGOR, D. Getting effective leadership in an industrial organization. *J. consult. Psychol*, 1944, **8,** 55–63.

SANFORD, F. H. *Authoritarianism and leadership.* Philadelphia: Institute for Research in Human Relations, 1950.

TANNENBAUM, A., & ALLPORT, F. H. Personality structure and group structure: An interpretative study of their relationship through an event-structure hypothesis. *J. abnorm. soc. Psychol.*, 1956, **53,** 272–280.

(Received July 25, 1958)

FAMILIAL ATTITUDES IN PARANOID SCHIZOPHRENICS AND NORMALS FROM TWO SOCIOECONOMIC CLASSES[1]

ROBERT C. LANE

North Shore Neuropsychiatric Center, Roslyn, New York

AND JEROME L. SINGER

Teachers College, Columbia University

THE results of a number of research projects indicate that there exists a relationship between personality development (including patterns of psychopathological behavior) and subcultural groupings in American society. From a theoretical standpoint these findings imply that cultural variations in familial organization play a critical role in personality differentiation. In practical terms these studies suggest numerous considerations concerning diagnosis and management of psychopathological conditions in persons of differing subcultural group origin, even where these groups live in close physical proximity (Hollingshead & Redlich, 1958; Opler, 1956; Srole & Langner, 1957).

The present study represents an attempt to investigate some of the differences to be found in attitudes toward parental figures of schizophrenic and normal individuals drawn from populations of middle and lower socioeconomic status. Considerable theory, clinical evidence, and some formal research data all suggest that a schizophrenic personality development arises from fairly specific patterns of interpersonal relationship in the family constellation of the growing child (Jenkins, 1950; Prout & White, 1950; Reichard & Tillman, 1950; & Terry & Rennie, 1938). If this is indeed the case, then certain variations in the descriptions of parental figures and of the family constellation should be observable when schizophrenic subjects are compared with "normals." To the extent that higher or lower socioeconomic status also implies somewhat diverse configurations of family relationships, as well as different child-rearing patterns, one may also anticipate differences in the attitudes toward parents of adults who come from different social classes (Auld, 1952; Grey, 1949; Needleman, 1951; & Singer, 1954). Finally, since it is reasonable to suppose that "schizophrenogenic" (Gerard & Siegal, 1950; Needleman, 1951; Prout & White, 1950; Reichard & Tillman, 1950) family constellations exist within the framework of specific subcultural influences upon parental interaction and child-rearing practices, a differential effect of social class status upon the familial attitudes of the two nosological groups may be expected. Thus, there are three general classes of hypotheses tested in the present investigation: personality differences in attitudes toward parental figures (normal–schizophrenic dichotomy), class differences (middle-lower-class dichotomy), and the interaction of these factors in determining attitudes toward parents.

Evidence of subcultural group influences upon familial attitudes as projected into Thematic Apperception Test stories has been provided in studies by Mitchell (1950), Henry (1947), McArthur (1955), and Singer (1954). In the last-named study, both psychopathological and socioeconomic status variables were employed. Hypotheses concerning the presence or absence of familial figures and the characterization of familial situations as "cold or rejecting" were tested. Singer concluded, "Although schizophrenics generally manifest disturbed intrafamilial relations, their sociocultural backgrounds appear to modify their familial experience patterns." Because of the problems involved in replication of such qualitative analy-

[1] A portion of this paper was read at the meetings of the American Psychological Association, New York, 1954. Some of the material in this paper represents an abridgement of the senior author's dissertation submitted in partial fulfillment of requirements for the Ph.D. at New York University. The authors wish to acknowledge the valuable assistance of Isidor Chein, Jacob Cohen, and Harold Wilensky, who made methodological contributions to various aspects of the investigation; to Seymour Klebanoff, Montrose VAH, Robert Morrow, Kingsbridge VAH, and Aaron W. Bortin, North Shore NP Center for administrative facilitation and services as judges, and to Murray Stern and Joseph Sacks, who also participated as judges. Nancy Perez executed the drawings for the FAS.

ses, more extensive objective methods were devised to study the interaction of personality and socioeconomic class factors on familial attitudes.

DERIVATION OF HYPOTHESES

Normal–Schizophrenic Differences

Available studies with reasonable controls have suggested a general picture of the schizophrenic family constellation (Gerard & Siegal, 1950; Jenkins, 1950; Lidz & Lidz, 1949; Needleman, 1951; Prout & White, 1950; Reichard & Tillman, 1950; Terry & Rennie, 1938). The schizophrenic child appears to have been reared in a setting in which the mother's role in the family is dominant. The father apparently plays a decidedly minimal part in domestic interaction and thereby becomes less available to the child for purposes of identification and imitation. The mother tends to be described as rejecting, either overtly by neglect and early deprivation of affection and nurturance, or covertly through imposition of excessive restriction upon the motor and emotional expression of the child. Accordingly, interpersonal relationships are charged with friction and the child emerges from the familial experience overdependent, unable to express hostility, fearful of punishment and rejection, plagued by inferiority feelings, unable to assume adult responsibilities. Based upon this general picture of the parent-child relationships of schizophrenics, a general hypothesis and a series of subhypotheses were established to test in this investigation:

1. The schizophrenic and normal groups differ in attitude towards family figures and relationships, with the schizophrenics manifesting less "healthy" attitudes (as defined below) than the normal subjects.

(*a*) The schizophrenics reveal stronger feelings of rejection by their parents.

(*b*) The schizophrenics indicate greater control by parents.

(*c*) The schizophrenics reveal greater dependence upon parents.

(*d*) The schizophrenics manifest evidence of a greater problem regarding the expression of hostility (as shown by either more repression or denial of hostility or by more direct hostility).

(*e*) The schizophrenics place a greater value on being a "good boy," a more conventional submissive type of personality.

(*f*) The schizophrenics describe a greater degree of maternal overprotection.

(*g*) The schizophrenics reveal more evidence of Oedipal conflict.

(*h*) The schizophrenics describe their parents as more unhappy with each other or more ill-suited for each other.

Socioeconomic Class Differences

The second general hypothesis which deals with the middle-lower-class dichotomy stems directly from the work of Mitchell (1950) and Singer (1954), whose results suggest that different patterns of child-rearing and family organization have significant effects upon familial attitudes of subjects from differing socioeconomic levels. It should be noted that the class groups have been selected as two clearly distinguishable *relative* positions in the socioeconomic hierarchy. There is no implication of two distinct social classes. The primary consideration of the study was not to provide a comprehensive description of familial attitudes for two social classes, but rather to demonstrate the need to take cognizance of the relationship between family attitudes and socioeconomic class in any subsequent research on family constellations and psychopathology. The following general and specific hypotheses are related to the variable of socioeconomic status:

2. The two socioeconomic groups differ in familial attitudes in the following variables:

(*a*) The middle-class subjects describe parental figures as more ambitious and concerned with their achievement than do lower-class subjects.

(*b*) The middle-class subjects are less able to express hostility.

(*c*) The middle-class subjects describe their mothers as more overprotective.

(*d*) The lower-class subjects describe their parents as using more direct forms of control.

(*e*) The middle-class subjects reveal a different relationship with father figures.

(*f*) The lower-class group manifest less evidence of Oedipal conflict.

(*g*) The lower-class subjects reveal a greater degree of interparental friction.

(*h*) The lower-class subjects indicate greater independence.

Socioeconomic–Personality Interaction

Equally important in this investigation is the question as to whether the above hypothesized normal–schizophrenic differences hold when comparisons are made for subjects at different socioeconomic levels. Only a few real cues are available from the literature relative to this point. Some evidence indicates a greater tendency towards certain signs of emotional instability, judging from personality-test responses at least, in the lower-class socioeconomic groups (Auld, 1952). It seems likely that a person with a lower-class background encounters greater over-all frustration and deprivation throughout many areas of life, and this is probably reflected by reports of a higher incidence of mental illness in the bottom rungs of the socioeconomic ladder (Grey, 1949; Hollingshead & Redlich, 1958). Jenkins' (1950) work on frustration suggests that economic, educational, and nutritional as well as psychological (conflict in familial relationships) deprivations can lead to serious maladjustment. It is plausible to infer that the higher socioeconomic groups have been exposed to fewer of these generally frustrating experiences; and so one could anticipate that familial attitudes per se would contribute more to the genesis of schizophrenia in the middle-class than in the lower-class schizophrenics. The third general hypothesis relates, therefore, to the interaction of the socioeconomic and normal–pathological dichotomy:

3. Familial attitudes within the lower socioeconomic class groups bear less relationship to the normal–schizophrenic diagnosis than do the familial attitudes of middle-class subjects.

In the operational terms of this study, therefore differences, between lower-class normals and schizophrenics with reference to familial attitudes were expected to prove smaller than the differences in the familial attitudes of middle-class normals and schizophrenics.

METHOD

Subjects

The *Ss* for this study consisted of 48 paranoid schizophrenic unmarried male patients, of whose number half were from a clearly middle socioeconomic background and half from a lower socioeconomic background, and 48 "normal" unmarried male patients hospitalized for medical and surgical reasons who were also divided into equal groups by socioeconomic status. The schizophrenic sample was chosen from patients hospitalized at the Franklin D. Roosevelt VA Hospital, while the normals were hospitalized at the Kingsbridge VA Hospital. Only paranoid schizophrenics were employed, since the symptomatology in this group is apparent and yet such patients tend to be in sufficient contact to perform cooperatively on tests. All 48 schizophrenics had a history of a clear break with reality, hallucinations, and delusions, and each had been formally diagnosed by the hospital staff. Only clearly literate and cooperative patients were selected, with no history of convulsive therapy for at least 60 days prior to testing.

The control *Ss* were chosen from medical-surgical wards. They were convalescing from conditions such as operations for pilonidal cysts or hernia and from limb burns or fractures. Five criteria were used to determine eligibility for the control group. These were: the opinion of ward medical personnel; the opinion of the examiner based on *S*'s attitude and integrity; determinable neuropsychiatric history; history of felonious conduct; and, finally, score on the Cornell Index. A Cornell Index score of 11 or more—or the occurrence of a critical response to one "stop" question—brought about the elimination of the control *S* from the sample.

The schizophrenic and control *Ss* did not differ significantly in age (mean 27.5), religion (16 Catholics and 8 Protestants in each of the four groups), number of siblings (no "only" children were used), birth order, place of birth, intactness of family group through *Ss*' first eight years of life, ages of living parents, and years of education (mean 11.6).

Establishment of socioeconomic status was based upon both economic and social information with greatest emphasis on paternal status. Factors considered as criteria included the level of formal education of the *S* and his father (whenever available); father's occupation in *S*'s childhood and *S*'s occupation; *S*'s evaluation of the type of neighborhood in which he lived during his formative years; family ownership of real estate; and history of public relief or assistance. Father's income did not prove to be a feasible criterion. Ratings by independent judges employing these criteria served as the basis for assignment of an *S* to a given socioeconomic class grouping. Overlapping of the two class groups on any individual criterion was minimal or arbitrarily avoided so that there was a sharp distinction in socioeconomic status with no borderline cases employed.

Test Battery

The test battery consisted of two procedures, the Elias Family Opinion Survey (Elias, 1952), designated the EFOS, and a projective technique devised specifically for use in this study, the Family Attitude Scale, designated the FAS. The EFOS as used here consists of 157 incomplete statements concerning family relationships in which *S* is expected to fill in "always," "often," "sometimes," "rarely," "never," each carrying a weighting from one to five. Examples include: "Mothers _____ love their children equally," "Fathers are _____ afraid of their wives," and "Parents _____ force their children to eat." The items are grouped into subscales such as: Attitudes towards Mothers;

Attitudes towards Fathers; Oedipal; Struggle for Independence; Interparental Friction–Harmony; and Rejection of Child. Details of test construction and validity have been presented by Elias. This procedure was chosen because it is a readily available objective device dealing with specific items relevant to the hypotheses of the study. In an effort to enhance cooperation and comprehension of instructions, the questions were presented individually on 3 x 5 cards, and *Ss* were instructed to place each card in one of five large, plainly marked boxes depending on his choice of the appropriate term, "always" through "never."

To supplement the EFOS, it was thought desirable to construct a suitable projective test which relied less upon the printed word as basic stimulus material. Since suggestive results in the field of social class or normal–schizophrenic differences in familial attitudes had been obtained by Mitchell (1950), Needleman (1951), and Singer (1954), using TAT pictures involving family scenes, it was decided to elaborate on this approach. After extensive preliminary exploration (using patients not included in this study), 11 pictures especially drawn for this research were deemed suitable for use. These included simple line drawings such as "Young boy watching father shave," "Father, mother, and son at breakfast table," "Child seated with toys while mother prepares to leave house," and "Mother combing child's hair."

A total of 202 statements, averaging approximately 18 per picture, were eventually chosen on the basis of prior independent evaluation by judges, first for relevance to a particular theoretical dimension and, secondly, for appropriateness to the specific drawing. These statements included items such as: (for Family Breakfast scene) "The boy has pretty definite likes and dislikes about food and lets his parents know"; "The boy accidentally spills his soup on his father's newspaper"; "His mother nags at his father all the time"; "His father is quick to use his hands on the boy"; "The boy wishes his father would go away so that he can take care of his mother." The statements were all chosen as pertinent to one of the 10 specific categories relevant to the experimental hypotheses. A list of these variables is presented in Table 2.

The experimental instructions developed after preliminary trials called for *Ss* to inspect each picture and then to read the statements which were printed individually on cards. The *S* was instructed to assign the statement to one of five boxes labeled Definitely Possible; Possible; Undecided; Not Possible; Definitely not Possible; depending on how he thought the child in the picture "really feels." It was hoped to obtain in this fashion a relatively objective score based upon a projected impression of the respondent. Scoring was on a linear continuum ranging from Definitely Possible, weighted five or one, to Definitely not Possible which was weighted as one or five. Both EFOS and FAS thus employed five-point scales with higher scores generally indicating a greater tendency to perceive problem situations or to respond to the family picture as representing distortions in normal family relationships.

The reliability of the EFOS and FAS for the sample employed in this study was ascertained by the use of Cronbach's Alpha Coefficient (1951). For the EFOS reliability coefficients for the nine subscales ranged from .69 to .94, with only two scales (Family Inferiority and Struggle for Independence) below .80. For the 10 FAS categories, reliability coefficients ranged from .51 (Independence) to .93 (Interparental Friction). Only three of the subscales showed coefficients below .75. These categories or item groups which could be considered parallel for EFOS and FAS were intercorrelated to establish the extent of relationship of the two measurement devices. Correlations between eight EFOS scales and various appropriate categories or subcategories of FAS ranged between .47 and .64 for the two scales when administered to the total sample. There were no significant differences in the correlations of the two procedures when the schizophrenic and control groups were considered separately.[1]

RESULTS

The data were evaluated by means of an analysis of variance in a 2×2 factorial design. The source variables were Personality (schizophrenic–normal dichotomy), Socioeconomic status (middle- and lower-class differences), and Interaction of these source variables. All significant F ratios were further analysed by t tests to establish the relationships among means of the four subgroups. The means, standard errors of the difference between group means, and F ratios for the four groups on EFOS and FAS are presented in Tables 1 and 2 respectively. These findings appear best evaluated in relation to the three major hypotheses of the study.

The "Personality" Hypothesis: Normal–Schizophrenic Dichotomy

The first general hypothesis and its subhypotheses involved comparison of the schizophrenics as a group with normals as a group. Inspection of Tables 1 and 2 indicates that, while four EFOS subscales and 18 FAS major and subcategories show significant F ratios and significant differences between normals and schizophrenics, most of these results derive chiefly from the differences between middle-class normals and the other three groups. The means of the Middle-Class Normal *Ss* are lower than those of the other three groups on almost all of the FAS variables, while those of the Middle-Class Schizophrenics tend to be

[1] Reliability and intercorrelations, as well as other relevant data, are available from University Microfilms, Ann Arbor, Michigan, Doctoral Dissertation Series, Publication No. 12,363.

TABLE 1

MEANS, STANDARD ERROR OF THE DIFFERENCE BETWEEN MEANS, AND *F* RATIOS FOR THE EFOS SUBSCALE

Scale	Mean MCN	Mean LCN	Mean MCS	Mean LCS	*F* Ratios			
					σ_D	P	S	I
A—Attitudes toward Mother	64.97	70.18	79.98	73.60	5.12	6.50*	.03	2.57
B—Attitudes toward Father	70.51	74.43	83.15	76.14	5.00	4.12*	.19	2.39
D—Oedipal	13.93	14.93	17.18	15.26	1.41	3.24	.21	2.19
E—Struggle for Independence	22.31	23.56	25.40	22.14	1.47	.59	.89	4.51*
F—Parent-Child Friction–Harmony	53.08	55.84	60.21	54.29	3.39	1.36	.44	3.27
G—Interparental Friction–Harmony	32.02	33.40	35.53	34.40	2.42	1.73	.01	.53
H—Family Inferiority–Superiority	28.77	31.69	33.90	31.07	1.61	3.89	.00	6.40*
I—Rejection of Child	34.11	37.95	40.82	39.74	2.31	6.76*	.70	2.26
J—Parental Qualities	27.23	29.36	34.65	31.69	2.20	9.89**	.07	2.68

* Significant at .05 level.

** Significant at .01 level.

Key: σ_D—Standard Error of the difference between means.
P—Variance attributable to "Personality."
S—Variance attributable to Socioeconomic status.
I—Variance attributable to Interaction.

highest. This finding clearly limits the generality of the "personality" hypothesis since the differences between the total schizophrenic and normal groups are chiefly attributable to the disparity between the two Middle-Class samples. If we consider each subhypothesis concerning presumptive differences between schizophrenics and normals in turn, the data yield the clear impression that while most of these hypotheses are supported, they hold frequently enough for the normal–schizophrenic dichotomy in only one social class group.

Hypothesis 1(*a*), for example, deals with feelings of rejection by parents, and it was tested by results on the EFOS Rejection of Child scale (EFOS Scale I) and by the FAS Total Rejection and Rejection by Mother and Rejection by Father variables (FAS Variable X). Inspection of Table 1 reveals a significant *F* ratio for Scale I, both normal groups yielding lower means than both schizophrenic samples. For the FAS Variable X, Table 2 reveals that while normals generally indicate significantly less total rejection and maternal rejection, Socioeconomic status and the Interaction also contributed significantly to the variance of this dimension.

Hypothesis 1(*b*) deals with presumed greater indications of parental control on the part of the schizophrenics. For EFOS it was felt that Scale E (Struggle for Independence) might afford a test, while for FAS the total Control and its subcategories (Variables II and IIa, IIb, IIc, and IId) appeared relevant. No clear-cut personality differences emerged on the EFOS scale, but the interaction between the Personality and Socioeconomic source variables proved to be significant, since the Middle-Class groups differed markedly while the Lower-Class groups clustered together. For the FAS scale the schizophrenics expressed significantly more Parental Control by Stress on Duty and Obligation but, again, on this and on most of the other control variables significant interaction effects emerged.

Hypothesis 1(*c*) (Dependence) received support from the FAS (no relevant EFOS scale was available here). Schizophrenics clearly show greater dependence on parents than normals in these items (Variable III) although again socioeconomic factors also show significance. The schizophrenics particularly reveal great dependence on mother, and this proves to be one of the most clear-cut results of the study.

With respect to Hypothesis 1(*d*) (Hostility), the schizophrenics demonstrate significantly more hostility toward parents on the Parental Qualities scale of the EFOS as well as on the Total Direct Hostility and Direct Hostility to Mother categories of the FAS. No significant findings for indirect hostility emerged.

For Hypothesis 1(*e*) (Good Boy) schizophrenics showed significantly more emphasis on this conventional-submissive pattern than

TABLE 2

MEANS, STANDARD ERROR OF THE DIFFERENCE BETWEEN MEANS, AND *F* RATIOS FOR THE FAS VARIABLES

Variable	Mean MCN	Mean LCN	Mean MCS	Mean LCS	*F* Ratios			
					σ_D	P	S	I
I. Ambitious Parents	37.15	39.36	37.74	38.53	1.48	.00	2.05	.46
Mother	18.39	19.39	18.68	18.72	.94	.00	.57	.57
Father	18.77	19.97	19.06	19.81	.73	.00	3.62	.31
II. Control	116.59	134.07	134.77	130.06	6.18	2.63	2.13	6.45*
a. By Physical Punishment	27.52	31.94	31.28	28.44	2.15	.00	.27	5.69*
Mother	13.09	15.18	15.60	14.34	1.21	.91	.23	3.87
Father	14.43	16.76	15.68	14.09	1.18	.72	.18	5.57*
b. By Restriction	28.77	34.32	34.40	33.44	1.82	3.44	3.19	6.40*
Mother	14.43	17.89	17.84	17.35	1.10	3.38	3.58	6.55*
Father	14.34	16.43	16.55	16.10	.90	2.14	1.63	3.98*
c. By a stress on Duty and Obligation Mother only	27.02	32.15	34.15	32.98	1.71	10.84**	2.68	6.79*
d. By Bribery or Promises Mother Only	33.28	35.65	34.94	35.19	1.68	.24	1.22	.83
III. Dependence	29.52	33.57	34.07	34.94	1.61	6.80*	4.70*	1.94
On Mother	13.05	16.14	17.22	17.26	.93	16.14**	5.57*	5.38*
On Father	16.47	17.43	16.85	17.68	.88	.22	2.11	.11
IV. Hostility	116.63	129.98	130.48	131.23	5.14	4.32*	3.76	3.00
a. Direct	24.19	28.90	31.07	28.81	2.11	5.20*	.68	5.48*
To Mother	12.09	14.34	16.35	14.18	1.24	5.46*	.00	6.39*
To Father	12.09	14.55	14.72	14.64	1.03	3.44	2.66	3.05
b. Indirect	39.87	47.91	47.29	46.16	2.96	1.84	2.73	4.80*
To Mother	28.40	33.49	32.90	32.32	2.13	1.24	2.26	3.55
To Father	11.47	14.43	14.39	13.84	1.01	2.69	2.85	5.95*
c. Denial and Repression	52.58	53.17	52.13	56.25	1.81	1.05	3.39	1.94
Denial	34.74	35.78	34.57	37.95	1.32	1.15	5.58*	1.58
Repression	17.85	17.39	17.56	18.31	.85	.23	.00	1.04
V. Idealization Pattern	56.34	57.59	55.67	63.93	2.25	3.18	8.96**	4.87*
a. "Good Boy"	20.93	21.43	22.02	25.19	1.05	10.64**	6.08*	3.27
b. Ideal Mother	19.43	19.97	17.68	21.64	.98	.00	10.60**	6.41*
c. Ideal Father	15.97	16.18	15.97	17.10	.68	.91	2.01	.91
VI. Independence	52.75	55.00	55.25	56.00	1.58	2.46	1.80	.43
a. Of Parents	27.86	29.77	29.57	30.61	1.05	2.87	3.93	.38
b. Of Mother	17.97	18.35	18.31	18.10	.90	.00	.00	.21
VII. Interparental Friction	61.22	75.64	76.98	75.89	4.90	5.33*	3.70	5.01*
a. One Parent Strong One Parent Weak	25.27	30.90	31.94	31.07	1.85	6.81*	3.28	6.18*
Mother Strong, Father Weak	12.01	14.93	15.68	15.18	1.02	7.36**	2.80	5.68*
Father Strong, Mother Weak	13.26	15.97	16.26	15.89	1.10	3.49	2.19	3.90
b. Ill-suited Marriage	11.84	15.18	14.93	15.10	1.25	2.88	3.94	3.19
c. One Parent Rejects Other	24.10	29.57	30.11	29.73	2.20	3.84	2.61	3.43
Mother Rejects Father	11.38	14.55	15.26	14.85	1.27	5.41*	2.34	4.06*
Father Rejects Mother	12.72	15.01	14.85	14.89	1.12	1.61	2.21	2.01
d. Ill-suited Marriage plus One Parent Rejects Other	35.95	44.74	45.04	44.83	3.32	3.82	3.35	3.68
VIII. Oedipal Relationships	36.65	43.62	46.08	44.54	2.76	7.01**	1.93	4.75*
Son-Mother Attachment	25.44	29.82	30.11	30.90	1.79	5.16*	4.15*	2.00
Son-Father Attachment	11.22	13.80	15.97	13.64	1.14	8.05**	.00	9.33**
IX. Overprotection	54.42	59.55	57.55	62.09	2.01	3.99*	11.59**	.02
By Mother	38.24	41.45	39.87	42.99	1.41	2.49	10.00**	.00
By Father	16.18	18.10	17.68	19.10	.82	4.73*	8.33**	.12
X. Rejection	31.82	40.99	42.12	41.57	2.71	8.06**	5.07*	6.44*
By Mother	20.35	26.69	27.94	27.06	1.89	8.89**	4.19*	7.30**
By Father	11.47	14.30	14.18	14.51	1.05	3.90	4.58*	2.90

* Significant at .05 level.
** Significant at .01 level.
Key: σ_D—Standard Error of the difference between means.
P—Variance attributable to "Personality."
S—Variance attributable to Socioeconomic status.
I—Variance attributable to Interaction.

did normals in the relevant FAS category (Variable Va). Again socioeconomic factors played a contributing role here. The Overprotection category (Variable IX) as measured by FAS (no EFOS scale was suitable), while revealing a significant personality effect, was much more strikingly influenced by socioeconomic status. The chief schizophrenic–normal difference here lay in the schizophrenics' greater emphasis on "overprotection by father," interestingly enough, while "overprotection by mother" proved to be chiefly a socioeconomic class preoccupation.

The Oedipal attachment hypothesis (Hypothesis 1[*g*]) failed to receive support in the relevant EFOS Subscale D, although the trend of the results was in the direction hypothesized. Stronger support came from the FAS (Variable VIII) with indications that schizophrenics generally manifest greater mother-son and father-son attachments with erotic, overly possessive overtones. Again socioeconomic factors also yield significant results.

Finally, Hypothesis 1(*h*) (Ill-suited Parents) receives support from the FAS data in particular. The Total Interparental Friction, (Variable VII) and One Parent Strong, One Weak, Mother Strong, Father Weak, and Mother Rejects Father categories all yielded significant results for the personality variable. The schizophrenics plainly describe parental figures as showing greater conflict and particularly emphasize a family constellation in which the mother is dominant and critical of the father. Here again significant interaction findings demonstrate that socioeconomic status is a contributory factor.

In summary, then it is clear that while the schizophrenics as a group do show a general tendency to differ from normals by describing familial patterns as involving greater rejection, maternal dependence, direct hostility, Oedipal attachments, and parental friction, these results cannot be considered independently of socioeconomic status. Any inference predicated on the supposed existence of a "schizophrenogenic" family constellation is untenable unless cognizance is also taken of sociocultural factors influencing the family organization or child-rearing patterns

Socioeconomic Class Hypothesis: Middle-Lower-Class Dichotomy

In connection with the second general hypothesis concerning socioeconomic class differences in family attitudes, no significant results emerged on EFOS. It should be noted, however, that the middle-class *S*s clustered at the extremes on this test; the middle-class normals scoring the lowest means (healthiest reactions) and the middle-class schizophrenics the highest means. There were 13 significant findings on the FAS in which socioeconomic class factors played a role. In all but three instances "personality" and "interaction" source variables also contributed significant variance. On an over-all basis there seems to be at least partial support for the general hypothesis of socioeconomic class differences in familial attitudes.

The more specific hypotheses yielded varied results. Hypothesis 2(*a*) (Ambitious Parents) receives no support from the data. The data bearing on Hypothesis 2(*b*) (Hostility shows no support from EFOS, and the FAS results are contrary to the stated hypothesis. Denial of Hostility (Variable IVc) revealed that lower-class *S*s were less able to express hostility about parents than the middle-class groups, which is the reverse of the hypothesis. Hypothesis 2(*c*) (Overprotective Mother) was not tested by EFOS. The results for FAS are highly significant for socioeconomic status but in a direction contrary to the hypothesis. The lower-class *S*s responded to the pictures by ascribing considerably more overprotection to mothers than the middle-class *S*s, and this is again the reverse of the hypothesis.

While no significant results support Hypothesis 2(*d*) (Direct Control), trends in support of Hypothesis 2(*g*) (Interparental Friction) and 2(*h*) (Independence of Parents) are discernible from the FAS data. The lower-class *S*s more frequently see parents as involved in an Ill-Suited Marriage and also indicate greater independence than do middle-class *S*s. Of special interest are the indications that at least on the FAS the lower-class *S*s show a significantly greater "son-mother attachment" again in direct opposition to Hypothesis 2(*f*) (Oedipal Conflict). Finally, inspection of variables bearing on Hypothesis 2(*e*) (Relationship with Father) indicates a strikingly different pattern

of response for the class groups as well as for the personality dichotomy. The lower-class *S*s show greater emphasis on overprotection, idealization, dependence on fathers, and the need to be a "good boy" at the same time as strong feelings of Rejection by Father (Variable X). These findings suggest that lower-class *S*s may wish for a closer although possibly less "healthy" relationship with their fathers.

It seems possible to conclude that, while socioeconomic status unquestionably appears to be contributing to differences in these *S*s attitudes toward parental figures, verification of the specific subhypotheses is by no means clear-cut. The fact that data relevant to two subhypotheses showed significant results in a direction opposite to that originally hypothesized supports the general indication of the significant role of socioeconomic status while at the same time raising important questions concerning the original interpretation of the nature of family organization of the two subcultural groups.

The Interaction Hypothesis: Socioeconomic–Personality Interaction

Inspection of Table 1 indicates that two of the EFOS scales show significant interaction findings. For these two subscales (Struggle for Independence and Family Inferiority–Superiority) significant differences emerged only between the middle-class groups. The middle-class normals show significantly less struggle for independence and fewer problems related to family inferiority–superiority than the middle-class schizophrenic group. Although only two of the interaction *F* ratios in EFOS are significant, the mean of the middle-class normal group is lowest on eight of nine subscales, reflecting a tendency towards relatively "healthier" attitudes in this group. Differences between the middle-class normal and middle-class schizophrenic groups were consistently larger than the differences between the lower-class normal and lower-class schizophrenic groups, thus affording support to the hypothesis that familial frustration—of and by itself—contributes more to the development of a schizophrenic personality in the middle-class than it does in the lower-class.

The analysis of significant interaction differences on the FAS by *t* tests reveals that when compared with the middle-class normal group, the other three groups cluster together in their mean scores. The middle-class normal group differs significantly from the other three groups on 17 of the 50 FAS variables. On the other variables the middle-class normal group differed significantly from the middle-class schizophrenic group, but not from the lower-class groups. All differences were in the direction of lower mean scores for the middle-class normal group. On all of these 17 dimensions, the middle-class groups differed significantly from each other, while the lower-class *S*s did not. On only two subscales did the lower-class groups differ significantly, the lower-class normal group expressing greater "physical punishment by father" and less "total idealization pattern" than the lower-class schizophrenic group. These instances are the only exceptions to the hypothesis.

Although three groups clustered, there were some significant differences among them. Relevant significant *t* test findings pertaining to differences between the middle-class schizophrenics and the lower-class groups on both tests include manifestations of less "idealization of mother," greater "struggle for independence" (EFOS), less "total idealization pattern," and greater "Oedipal son-father attachment" for the middle-class schizophrenics. The lower-class schizophrenic group differed significantly from both middle-class groups on "idealization of mother" and from the other three groups on "total idealization." A significant finding worth noting is that the lower-class normal group showed significantly greater "total physical punishment" than the middle-class normal group.

For the 59 variables considered (9 on EFOS and 50 on FAS), there are 24 instances of significant interaction findings. In 20 of these instances, there were significant differences between the middle-class groups but not between the two lower-class groups. If direction alone and not significance is considered, for 49 of the 59 variables there are larger differences between the two middle-class groups than between the two lower-class groups. It would appear that there is certainly strong evidence in support of the interaction hypothesis as stated. This is particularly the case in the results on the projective-type FAS and less striking in EFOS.

DISCUSSION

In the attempt to formulate an interpretation of the findings of this study, the assumption has been made that familial attitudes are well established in early life. While their expression in adult life may well be colored by subsequent experience and the impact of psychopathology, these attitudes must have had real roots developed in childhood. From this standpoint, the most striking result of the study was an indication that socioeconomic class differences are indeed influential in contributing to the differences between schizophrenics and normals in their early attitudinal development. The majority of the significant interaction effects were accounted for by the difference between the middle-class normals and the other three groups. The middle-class normal *S*s indicated by their responses that they had experienced much less family friction, homelessness, control, hostility, and rejection than the other groups. The middle-class schizophrenics differed most clearly from the other groups by indicating greater feelings of rejection by parents, greater hostility toward parents, more struggle for independence, Oedipal involvement, and maternal control by stress on duty and obligation. It seems clear, taking all the findings into account, that the picture of the "schizophrenogenic" family constellation as described above applies chiefly to patients from the middle-class. The finding of a "reverse Oedipal pattern" with middle-class schizophrenics expressing more resentment of mothers and significantly greater "Oedipal attachment for father" than all other groups appears to support Freudian views of the homosexual aspects of paranoia. Yet this finding, too, remains a middle-class phenomenon. By contrast the lower-class *S*s, normal and schizophrenic together, present a picture of what is essentially a denial of negative parental influences and of almost an idealization of parents. This is particularly true of the schizophrenics who emphasize a "good boy" pattern.

Evaluation of the data also suggests greater conscious conflict on the part of the middle-class schizophrenics whose negative feelings and ambivalence concerning their parents seem to have burst forth to torment them. Lower-class *S*s, on the other hand, seem more likely to repress or deny childhood problems, striving to maintain an image of parents in more idealistic terms, although the lower-class normals appear to be more realistic in indicating a history of greater physical punishment. One may speculate, then, that the particular pathogenic factors operative in middle-class families are likely to lead to extensive ruminative conflicts, struggle for independence, and extensive self-analysis. The lower-class schizophrenic perhaps because he has been crushed by adversity or smothered by a particular guilt-ridden, overprotective mother, may be more inclined to repress conflict and quite possibly to act out his difficulties through symptoms or antisocial behavior with little conscious awareness of the extent of negative feeling.

In this connection, the findings of this study bear comparison with results obtained by Grey (1949) in an investigation of the relationship of social status to the psychological characteristics of a variety of hospitalized psychiatric patients. Grey found, on the basis of extensive interview with middle- and lower-class patients, that the former group tended to conform more to hospital rules and to have led more conventional lives but were more likely to *verbalize* aggressions and rejection of authority in the clinical setting. Middle-class patients "talked out hostile attitudes," while the lower-class patients "acted them out." Furthermore, the middle-class patients, while reporting that they had received more material assistance from families to a later age and while indicating that parents had participated more in their social and school activities, still verbalized more direct hostility to mothers in particular. Physical punishment, according to Grey, had been used less by middle-class patients' parents with greater stress on discipline through invocation of moral standards, a result that confirms the findings on control by stress on duty and obligation in this investigation.

The "idealization" pattern or denial of hostility to parents by the lower-class patients in this study is of special interest in view of Grey's findings that, while lower-class patients left home earlier and received less material and psychological attention from parents, they began contributing support to their parents earlier and continued it longer and with proportionately greater amounts of their income.

One wonders whether the "overprotection by mother" which is expressed most strongly

by the lower-class *S*s, normal and schizophrenic, may not represent a kind of sentimental wish-fulfilling reaction in this instance. It is as though, within the rigors of economic and social deprivation which characterizes lower-class life, the nurturing mother (who may of necessity have taken this role for only a relatively short time in the children's lives) becomes a symbol and focus of positive wishes.

The results of the present investigation have important implications for those theories which stress a psychogenic etiology in schizophrenia as well as for therapy and management of patients with varying socioeconomic backgrounds. The data certainly offer considerable support for the belief that schizophrenics develop in family constellations in which there is strong maternal domination, a relatively weak father, rejection of the child, control through guilt-provoking mechanisms, and apparently a tendency for the child to develop a spiteful reaction towards the controlling mother and an inverted Oedipal complex with homosexual implications. Yet this is chiefly a middle-class phenomenon in this investigation, and it is noteworthy that the clinical studies which have provided data on "schizophrenogenic" families seem to have been based largely upon populations of middle-class patients. Because of similar symptomatology it has been assumed that similar dynamic factors have operated for lower-class schizophrenics. Perhaps, as Jenkins (1950) suggests, extensive frustration may well have been the crucial factor, but the frustration may be in different areas, closely related to familial factors for the middle-class schizophrenic group and to environmental factors in the lower-class schizophrenic group. It is clear that any attempt at further formulation of a theory of the psychodynamics of schizophrenia must take into account the cultural pattern and its impact on family organization and child-rearing patterns. A similar conclusion is drawn by Hollingshead and Redlich (1958, p. 365) who note that "presses" are greater for the lower strata while internal "stresses" are more marked in the middle classes.

From the standpoint of practical handling of patients (ward management or therapeutic approach), the findings are also provocative. Singer and Opler (1956), for example, found that Irish and Italian schizophrenic patients differed in their cooperation with ward routines and aggressiveness in day-to-day behavior. Grey (1949) reported that middle-class psychiatric patients received more therapeutic attention from staff, more psychotherapy, and were viewed as prognostically more hopeful by the hospital professional staff, which naturally consisted of middle-class personnel. The present data suggest a reason for Grey's finding in addition to the probable staff identification with the middle-class patients. The lower-class schizophrenic patients with their greater "idealization patterns," greater need to be "good," greater "denial of hostility —a finding reported also by Tomkins (1952) in studies with the Picture Arrangement test— might not prove especially responsive to insight therapy which threatened their rigid defense systems. Group therapy might prove more advisable. Here self-expression might be easier with group support. For these lower-class schizophrenics to admit that they hate their mothers, if indeed they do, would be a *sin* leading to fantasied retribution. The fleeting awareness of such hateful thoughts may have led to their breakdowns and therefore re-repression may prove more effective. Actual physical outbursts or hostile episodes which characterize the histories of lower-class patients are often followed (sometimes with the assistance of shock treatment) by amnesia for the episodes and by subsequent denial of hostility and idealization of parents. On the other hand, middle-class schizophrenic patients with their greater overt hostility, greater verbal rebellion, lesser need to appear as "good boys," and introspective conflicts might prove more amenable to insight therapy.

The data of the present study can scarcely be considered sufficient to call for a decided change in patient management or assignment to treatment. Nevertheless, by pointing up the consistent differences in familial attitudes of schizophrenics and normals from differing socioeconomic levels, they raise quite important questions concerning the limitations of an over-all dynamic view of schizophrenics or other psychiatric patients. It seems certain that one must seriously consider the differential influences on personality development of various subcultural patterns of family organi-

zation for the fullest understanding of the psychogenesis of schizophrenia.

SUMMARY

The objective of this study was to determine whether certain differences in attitudes toward parental figures obtain between paranoid schizophrenics and normals and whether these differences continue to hold when comparisons are made at different socioeconomic levels. Three general hypotheses and a series of specific subhypotheses were offered. The first, the "personality" hypothesis, concerned the differences between normal and schizophrenic groups; the second, the "socioeconomic class" hypothesis, bore on differences in familial attitudes between subjects of middle and lower socioeconomic status; and the third, the "interaction" hypothesis, involved the prediction, based on a frustration theory of schizophrenia, that familial attitudes of middle-class normals and schizophrenics differ more markedly than do those of lower-class normals and schizophrenics.

Ninety-six white male veterans served as subjects. They were divided into four equal groups, middle- and lower-class paranoid schizophrenics and middle and lower class "normal" veterans hospitalized for minor medical and surgical care. Subjects were American born, single, Protestant or Catholic, between the ages of 20 and 34, with at least one sibling and with a history of intact family for at least the first eight years of life. Various controls were employed to insure clear contrasts between groups on socioeconomic status and psychopathology and to rule out extraneous factors. Socioeconomic status was established on the basis of paternal and subject education, occupation, type of neighborhood during formative years, property-ownership, receipt of public relief, etc.

The Elias Family Opinion Survey and a specially constructed Family Attitude Scale were administered. The latter (FAS) consisted of a number of pictures of family scenes with a number of statements whose relevance to the picture the subject was asked to rate along a five-point scale. These statements represented aspects of 10 variables bearing on familial attitudes which were scored from subjects' ratings.

Results indicated considerable support for the "personality" hypothesis with the qualification, however, that many of the significant differences between normals and schizophrenics were accompanied by significant socioeconomic or interaction findings. The schizophrenics generally showed greater maternal dependence, greater emphasis on "good boy" patterns and Oedipal attachment. They felt more "homeless" and experienced greater interparental friction and maternal rejection, viewing the mother as the dominant parent about whom family life pivots and who maintains control by stress on duty and obligation. The "social class" differences suggested that the lower-class subjects experienced greater idealization of the mother, dependence on her and yet greater independence of the family itself. They saw parents as more ill-suited and experienced greater rejection, particularly by father. More striking support was obtained, however, for the "interaction" hypothesis. On 49 of the 59 subscales and variables studied in this investigation (EFOS and FAS), the difference between the middle-class normals and schizophrenics was greater than the difference between lower-class normals and schizophrenics. The differences were largely based on the tendency of the middle-class schizophrenics to verbalize considerably more conflict and problem in family relationship, while the lower-class pattern seemed to stress denial of hostility and overidealization of the mother. Contrasts within the lower-class groups' attitudes toward familial figures were also obtained, but these were less frequent than those between normal and schizophrenic middle-class subjects. The results of the study were discussed in terms of the implications of the findings for the role of sociocultural and idiosyncratic factors in family constellations as sources of personality differences in later life. Implications for treatment and evaluation of schizophrenics as well as the need for consideration of socioeconomic factors in further research on psychopathological groups were also discussed.

REFERENCES

AULD, B. F., Influence of social class on personality test responses. *Psychol. Bull.*, 1952, **49**, 318–332.

CRONBACH, L. J. Coefficient alpha and the internal structure of tests. *Psychometrika*, 1951, **16**, 297–334.

ELIAS, G. A measure of "homelessness." *J. abnorm. soc. Psychol.*, 1952, **47**, 62–66.

GERARD, D. L., & SIEGAL, L. The family background of schizophrenia. *Psychiat. Quart.*, 1950, **24**, 47–73.

GREY, A. L. Relationships between social status and psychological characteristics of psychiatric patients. Unpublished doctoral dissertation, Univer. of Chicago, 1949.

HENRY, W. E. The thematic test technique in the study of culture. *Genet. psychol. Monogr.*, 1947, **35**, 3–135.

HOLLINGSHEAD, A. B., & REDLICH, F. C. *Social class and mental illness*. New York: Wiley, 1958.

JENKINS, R. L. Nature of the schizophrenic process: A working hypothesis for therapy. *AMA Arch. Neurol. Psychiat.*, 1950, **64**, 243–262.

LESHAN, L. L. Time orientation and social class. *J. abnorm. soc. Psychol.*, 1952, **47**, 589–592.

LIDZ, R. W., & LIDZ, T. The family environment of schizophrenic patients. *Amer. J. Psychiat.*, 1949, **106**, 332–345.

MCARTHUR, C. Personality differences between upper and middle classes. *J. abnorm. soc. Psychol.*, 1955, **50**, 247–254.

MITCHELL, H. E. Social class and race as factors affecting the role of the family in thematic apperception test stories of males. Unpublished doctoral dissertation, Univer. of Pennsylvania, 1950.

NEEDLEMAN, S. Ideational concepts of parental figures in paranoid schizophrenia. Unpublished doctoral dissertation, New York Univer., 1951.

OPLER, M. K. Cultural anthropology and social psychiatry. *Amer. J. Psychiat.*, 1956, **113**, 302–311.

OPLER, M. K., & SINGER, J. L. Ethnic differences in behavior and psychopathology: Italian and Irish. *Int. J. soc. Psychiat.*, 1956, **2**, 11–22.

PROUT, C. T., & WHITE, MARY A. A controlled study of personality relationships in mothers of schizophrenic male patients. *Amer. J. Psychiat.*, 1950, **107**, 251–256.

REICHARD, SUSAN, & TILLMAN, C. Patterns of parent-child relationships in schizophrenia. *Psychiat.*, 1950, **13**, 247–288.

SINGER, J. L. Projected familial attitudes as a function of socioeconomic status and psychopathology. *J. consult. Psychol.*, 1954, **18**, 99–104.

SINGER, J. L., & OPLER, M. K. Contrasting patterns of fantasy and motility in Irish and Italian schizophrenics. *J. abnorm. soc. Psychol.*, 1956, **53**, 42–47.

SROLE, L., & LANGNER, T. S. Treated and untreated mental disorders in the metropolis. Paper read at American Sociological Society, Washington, D. C., 1957.

TERRY, G. C., & RENNIE, T. A. C. Analysis of paregasia. *Nerv. ment. Dis. Monogr.*, 1938, No. 64.

TOMKINS, S. S. The Tomkins-Horn picture arrangement test. *Trans. N.Y. Acad. Sci.*, 1952, **15**, 46–50.

(Received July 28, 1958)

POLITICAL ORIENTATION AND THE PERCEPTUAL EFFECTS OF AN ANTICIPATED INTERACTION[1]

EDWARD E. JONES AND BOICE N. DAUGHERTY[2]
Duke University

Do "BIRDS of a feather flock together" or do "opposites attract"? The paradox suggested by these incompatible aphorisms has stimulated a large number of investigations into the roles of similarity and complementarity in social preference. It seems reasonable to assume that we like and enjoy others who share our beliefs and values. On the other hand, the assumption that strong and compatible dyads are formed when the strengths of one member complement the weaknesses of the other, or when the needs of one are gratified by the behavior of the other, also has a common sense appeal.

It seems, then, that a similarity hypothesis is appropriate for some of the phenomena of social preference, whereas a complementarity hypothesis is more pertinent for others. A fact of crucial significance would seem to be the nature of the particular personality attribute being considered in the similarity or complementarity relationship. As indicated in Lindzey and Borgatta's (1954) summary of sociometric studies, mutual friends in the setting of an on-going group tend to have similar patterns of interest, value, and attitude. As an illustration of this kind of study, Precker (1952) found a closer relationship between the self-ranked values of mutually chosen associates than of unrelated peers in a small college setting. He suggests that "valuings operate in the selection of associates. . . since they allow for a *universe of discourse*, an operational 'language,' which facilitates intercommunication and, thereby, interaction" (p. 406). A number of other factors, including the emotional support derived from consensus about the things which are "important," could also underly this relationship.

When one departs from the explicit domain of values and interests, however, to deal with other personality characteristics such as traits (Van Dyne, 1940) or behavior tendencies (Zimmer, 1956), the reported relations between attributes of chooser and chosen become slight and inconsistent. Many of the variables that might be identified in this broader sphere suggest the importance of complementarity as opposed to similarity. Winch (1952, 1955; Winch, Ktsanes, & Ktsanes, 1954) has been the most outspoken proponent of complementarity in selective association. Working at the level of personality need, and following H. A. Murray in the definition of this construct, Winch has presented data on 25 young married couples which give qualified support to the predicted relation between mate selection and need-complementarity. As Zimmer (1956) has noted, however, Winch does not indicate what safeguards were exercised against possible contamination of ratings by knowledge of the hypothesis and the subjects. Furthermore, Bowerman and Day (1956) found no support for either the similarity or complementarity hypothesis with dating couples in a different study.

The distinction between values on the one hand and needs on the other hand does not, then, dramatically resolve the paradox. Some values (e.g., political) seem to call less for a resonant similarity in the determination of social preference than others (e.g., esthetic values). And in the area of needs, a simple complementarity hypothesis seems unnecessarily restrictive. In reviewing Winch's argument, Rosow (1957) comments that "while dissimilarities may be a source of strength [and equilibrium in a marriage], many *similarities* of need might be equally compatible and equally functional. For example, similar status aspiration needs might strengthen a

[1] This study was conducted in the program of the Organization Research Group of the Institute for Research in Social Science at the University of North Carolina. The O.R.G. is supported by the Office of Naval Research [Nonr-855(04)]. The research was actually conducted at Duke University and was facilitated by grants to the senior author from the Duke University Research Council and The Ford Foundation.

Albert Carlin and Keith Davis both made substantial contributions to this research, the former as the second experimenter, the latter as data analyst.

[2] Now at the University of Kentucky.

marriage just as dissimilar ones may weaken it" (p. 229).

The basis for discriminating between attributes favoring similarity and those favoring complementarity appears to be subtle and complicated. Perhaps the direction toward a solution lies in an attempt to analyze the most likely behavioral consequences of a particular personal attribute in different interaction settings. Let us take, for example, the need to dominate. In some settings, where attempts at social control and influence are perhaps encouraged, ascendant people may respond more positively to submissive people who increase the likelihood of gratifying their needs. In other settings, however, the dominator may seek out consensual justification for his desire to dominate by associating with and responding positively to other dominators. Through a detailed analysis of what the situation constrains the actors to regard as appropriate, perhaps it will be possible to achieve more precise prediction of liking in relation to congruent and reciprocal personal attributes.

Jones and Thibaut (1958) have proposed that different *inferential sets* are aroused by interactions or anticipated interactions which have different goals as far as the actors are concerned. Different characteristics of a stimulus person become salient depending on whether the perceiving actor's goal is to influence, to exploit, to receive value support, or merely to receive a "liking" response from the stimulus person. The same behavior of the stimulus person may thus take on different values for the perceiver, and we should be able to notice changes in his perception and evaluation of the stimulus person as a function of the interaction context.

To summarize, then, predicting interpersonal choice and perceptual impression is a hazardous and difficult task at best. In making our prediction we must consider (*a*) the manifest attributes of the stimulus person, the object of potential choice or perception; (*b*) the values, attitudes, and interests of the perceiver *in relation to* the apparent values, etc. of the stimulus person; (*c*) the underlying needs and interpersonal orientations of the perceiver in relation to need-gratifying attributes of the stimulus person. Finally, we must attempt to specify (*d*) the interaction context in which the perception and choice behavior take place. What is the nature of the existing or anticipated relationship between the perceiver and the stimulus person, and how does this affect the other three interrelated variables?

In order to explore the operation of one set of these four classes of variables, a controlled experiment was designed. *S*s listened to tape-recorded interviews between the *E* and two stimulus persons, one of whom was manifestly political in his beliefs and orientation, and one of whom was manifestly esthetic. In two treatments, *S*s monitored the interviews under conditions of deception in which they were led to anticipate interaction with the stimulus person of their choice. In a control treatment, *S*s were told that they would hear taped sessions of interviews recorded with students in a previous year. Two measures of political orientation were obtained from each *S*: his score on the political scale of the Allport-Vernon-Lindzey Study of Values, and his score on Christie's Mach IV Scale.

The broad intent of the study was to show that the perception of a political stimulus person depends both on the degree to which the perceiver is politically oriented and on the interaction context of the perceptual act. In particular, it was predicted that politically oriented people should tend to attribute more positive characteristics to a politically minded stimulus person in the abstract than when there is anticipation of interaction. Thus, the similarity hypothesis should accommodate the results when no interaction is anticipated, and the complementarity hypothesis be more appropriate under the condition of anticipated interaction.

Since the goal of the anticipated interaction was assumed to be of obvious importance, two types of interaction settings were described to different *S*s. While one setting was "competitive" and the other "cooperative," both were designed to increase relative preference for the esthetic stimulus person—particularly by those *S*s with strong political orientations. A secondary hypothesis, to be described more fully below, predicted that the relationship between expressed preference for a partner and assignment of positive characteristics to

the partner would be more positive when a cooperative interaction was anticipated. Nevertheless, the main reason for including two types of interaction goals was to provide some of the features of experimental replication.

METHOD

Subjects

Exclusive of *S*s used in pretesting, 83 undergraduate male volunteers participated in the experiment proper. Of these, the data of 72 *S*s were included in the intensive analysis of perception of the stimulus persons. The remaining 11 *S*s were omitted because of rating errors or suspicious rejection of instructions in the conditions where interaction was supposed to be anticipated.

Measures of Individual Difference

The political orientation of each *S* was measured by two different scales, the Allport-Vernon-Lindzey (1951) Scale of Values (AVL) and the Christie Mach IV Scale, which were administered in a classroom setting prior to the *S*s' participation in the experiment. Two scales were used partly as a check on each other, but mainly because we felt that each measured a rather different facet of political orientation. While the two correlated in the neighborhood of .30 with each other in a large undergraduate sample, the political items on the AVL seem to be more concerned with interest in public affairs, prestige, and competition, than with a direct interest in influencing and controlling others. The Mach IV Scale reflects a different pattern of personal qualities including strong cynicism and attribution of selfish and ulterior motives to others. This pattern involves a flexible morality which underlies and justifies interpersonal manipulation.

The properties and form of the AVL are well known and need no further discussion here. The Mach IV Scale, on the other hand, is a relatively new and unexplored measure, although a description of item content and some material comparing various samples may be found in Christie and Merton (1958). The scale was devised by Christie partly from statements made in *The Prince* and *Discourses* by Machiavelli to provide a measure of individual differences in accepting the ethics of manipulating others, employing tactics such as flattery when it serves a purpose, and tailoring the means of social intercourse to the end of augmenting one's own power and status. The version of the Mach scale used in the present study (the fourth of seven which have been developed in attempting to refine and purify the scale) contained 20 Likert-type items, 10 worded in each direction, such as "There is no excuse for lying to someone else" (scored negatively) and "Anyone who completely trusts anyone else is asking for trouble" (scored positively). In spite of a superficial resemblance to the F scale (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950), the Mach scale has shown low and inconsistent correlations with F in samples of college students, lobbyists, and medical students. While research detailing the relevant properties of the various Mach scales is as yet unpublished, Mach IV has been carefully pretested for internal consistency, item intelligibility, etc.

The original plan was to separate the experimental sample into highs (who scored above the median on both Mach IV and AVL) and lows (below both medians) for placement in a factorial design. Since it proved difficult to procure an adequate sample by these criteria, it was decided to base the major analysis on correlations between each scale score and flattery of the two stimulus persons in each treatment. As will be apparent below, this decision turned out to be fortunate.

The mean score for all *S*s on the political scale of the AVL was 43.04; *SD*, 8.10. The distribution of scores was thus very similar to the norms for college males reported in Allport, Vernon and Lindzey (1951, p. 9). The mean score for all *S*s on the Mach IV Scale was 73.08; *SD*, 16.93. Comparison data are not available, but these figures are very close to the mean (73.97) and standard deviation (15.73) of the classroom population of 198 male students from which the experimental sample was drawn. *S*s were randomly assigned to experimental conditions, and the means and standard deviations for AVL-political and Mach IV scores, respectively, were almost identical from condition to condition.

Procedure

In all treatments, the experimental sessions involved "groups" of *S*s ranging in actual number from one to seven. In all cases a strenuous effort was made to convince each *S* that a number of other *S*s, whom he had not yet seen, were also participating. This deception was quite possible because as soon as *S* reported, he was escorted to one of 10 identical experimental booths and from that point until the end of the experiment he actually saw no one. The entire experiment was conducted by tape recordings transmitted by individual loudspeakers into each experimental booth.

As soon as the *S*s were settled, tape-recorded instructions were fed into each booth providing a rationale for the experiment. The instructions varied with each treatment in the following manner:

1. *Control Condition: S*s in this treatment were merely told that they were participating in an "experiment in interpersonal perception and evaluation" and that they would hear two previously recorded interviews between the experimenter (E_1) and students selected from the introductory psychology course the year before. It was explicitly stated that the stimulus persons would not be present at any time and that the *S*s would not meet either of them. The *S*s were instructed to take notes on the interviews and to rate each stimulus person on the forms provided.

2. *Anticipation of Competitive Interaction* (AIComp): *S*s in this treatment were told that "this is a study of some of the things which determine how two people interact with each other in a discussion. In the main and final part of the experiment some of you will get a chance to talk together in pairs about the personal qualities which make for effective leadership. We want to see which member of the pair is most effective in the discussion." These instructions were later elaborated by informing *S* that he would be independently evaluated

in terms of the soundness of his ideas and "your ability to convince your partner that your ideas are better than his." Prior to this elaboration, the *S*s were told that some of them would be interviewed by E_1 and the rest of the *S*s were to listen. *S*'s role was to be determined by the letter on the wall of his booth. In all cases, this letter designated that *S* was to be a "listener." General instructions were given summarizing the *S*'s task: while the interviews were in progress, the listeners were to make notes on their first impressions after both interviews were completed, *S* was to "complete some forms" on each stimulus person; finally, each listener would be brought together to interact with one of the interviewees, the particular one to be chosen by lot. This last instruction was actually countermanded by a second experimenter (E_2) who presumably gave special instructions to the listeners while E_1 was giving special instructions to the interviewees. This was accomplished by having E_1 mumble in the background while E_2 actually instructed the *S*s that they would be able to *choose* which stimulus person they preferred to interact with. The instructions were handled in this manner so that *S* would realize that his preference was determinative but also feel that neither stimulus person would be aware of his role in deciding on an interaction partner. In this way we hoped to make the choice more genuine and free from inhibitions or fears of reprisal.

3. *Anticipation of Cooperative Interaction* (AICoop): *S*s in this treatment received instructions identical to those given in AIComp except for the description of the interaction task. This was to be a game of answering questions about "poetry, music, and famous literary figures." The *S* and his partner would work as a cooperative team and were to come to agreement on a joint answer for each question asked.

It may appear to the reader that the particular kinds of interactions described have an arbitrary relationship to the experimental purpose. Actually, our intent was to lower the valence of the political stimulus person in two quite different ways in hopes that the politically oriented perceivers in particular would change their ratings in a negative direction. Thus to the political *S*, the competitive interaction should make the rather aggressive and power-oriented political stimulus person seem less attractive, and the cooperative interaction should make the knowledgeable esthetic stimulus person more attractive because of his value to the "team."

All *S*s filled out postexperimental questionnaires and were interviewed at length by the two *E*s. Most of the *S*s were clearly surprised to learn that the interactions would not take place, and several expressed disappointment that they would not see the stimulus persons. A few were clearly suspicious, as noted above, and their data are omitted from analysis.

Stimulus Persons

The interviews were prerecorded by two psychology graduate students on two separate tapes in order to counterbalance the sequence of presentation and to control for extraneous vocal cues. Thus Interviewee I was first on each tape, but read the esthetic script on one and the political script on the other. Interviewee II read the remaining script on each tape. Approximately half of the *S*s in a given treatment heard the interviews in one sequence, and half of them in the other. Since there were no obvious effects of sequence in any condition, the data were pooled for all the major analyses discussed below.

The scripts themselves were carefully constructed and several times revised during pretesting. They were composed to reflect, in terms of the local college setting, high political or high esthetic values in terms of Spranger's descriptions. In order to give some of the flavor of the interviews brief excerpts are compared below:

Political Stimulus Person	Esthetic Stimulus Person
I'll tell you one thing. . . . I have to admit it's important to know where a person comes from and what kind of organizations he belongs to . . . uh . . . well, I think you can tell a lot about a person from things like that . . . whether he's got a lot on the ball. . . . A lot on the ball? Uh . . . you know, he's a person you can respect . . . someone who knows where he's going and how to get there. . . . I guess I always admire people who handle themselves well with others . . . who know how to get things done without getting others angry.	I think it would be a sad world if everyone was an "All-American boy." Well, take for instance this guy I know in the dorm. . . . Now, some guys I know think he's kind of a sad sack. Sometimes he says the wrong thing, and he argues a lot about religion and philosophy and the books that he's read. I mean, we have a good time talking, sometimes, about the books we've read and the music we both like.
I've always wanted to be a leader in various things. I mean . . . uh . . . I've always enjoyed playing an active part in things. I think you get out of a group just what you put into it, you know . . . and I kind of enjoy making decisions . . . or helping to make them.	. . . I'm not much of a joiner except I've gone along with my roommate to several play reading sessions at Branson and things like that. . . . I like to do various things with other people when I think I can contribute something, like playing in a quartet or something like that. . . .
I've pretty much decided to major in political science—especially if I actually end up going to law school which is the way things look now.	. . . I don't have the vaguest idea what I'll end up doing. I don't know, I think it's silly to wonder about whether you're wasting your parents' money here if you're learning something. . . .
I admire Churchill. To me he's a great man . . . kind of a symbol of courage. And I admire Roosevelt not so much for what he did, so much, as the way he did it. . . .	Well, I kind of admire Sigmund Freud, although I don't know too much about him. . . . I've always admired Schweitzer . . . uh . . . he seems to have fulfilled himself and at the same time managed to help other people. . . .

Rating Scale

In all treatments, the *Ss* indicated after hearing both interviews which stimulus person they liked best (control condition) or which person they wished to interact with (AIComp and AICoop). While these responses were of interest, and the questions themselves were essential to support the experimental rationale, the choice data alone could not serve as a valid test of the main hypothesis since it was difficult to control for the absolute appeal of each stimulus person. The most relevant dependent variable was a 30-trait rating scale which could be scored for halo effect or flattery by an algebraic summation of positive and negative characteristics assigned to each stimulus person. Such an index had proved sensitive and useful in an unpublished study by Jones, Bruner, and Tagiuri, which clearly showed that people tend to flatter others who have similar values. A similar index also provided meaningful data in an earlier study by the senior author (Jones, 1954).

The scale itself consisted of 15 favorable adjectives (such as Sincere, Well-rounded, Good sense of values, Sympathetic, and Sociable) and 15 unfavorable adjectives (such as Misguided, Pretentious, Shallow, Mercenary, and Irritating). The *Ss*' task was to place these adjectives in a forced, approximately normal, distribution which could be scored from one to seven. The number of adjectives which could be assigned (from 1, "highly characteristic", to 7, "highly uncharacteristic") to each score were as follows: 1, 3, 5, 10, 5, 3, 1. The flattery score for each stimulus person was derived by reversing the score of the positive items and adding their subtotal to the subtotal of negative item scores. A completely neutral impression would thus receive a core of 120, and the actual scores ranged from 86 (relatively unfavorable) to 170 (relatively favorable).

RESULTS

Validation of Stimulus Differences and Treatment Conditions

Differences between stimulus persons. It would have been extremely difficult to create two stimulus persons who were evaluated with the same degree of favorability by the total sample of *Ss*. Actually, across all conditions, the esthetic stimulus person (SP) was flattered more than the political SP ($t = 3.98$, df 71, $p < .001$). This result could have been a function of the particular traits included in the rating scale, since actual preference for the esthetic person was not significantly greater than preference for the political person as a potential interaction partner.

Of greater relevance are the traits which actually discriminated significantly between the two SPs regardless of experimental condition. A list of these traits, with their probability values, is provided in Table 1. As intended in the construction of the two scripts, the political SP was seen as more conformist, forceful, sociable, and authoritarian than the esthetic SP. The esthetic SP, on the other hand, was seen as more imaginative, creative, intellectual, and sympathetic as intended. The other reported differences are generally consistent with these examples. It may be concluded, then, that we were quite successful in conveying the intended differences via the interviews.

TABLE 1

TRAITS WHICH DISCRIMINATE BETWEEN POLITICAL AND ESTHETIC SPS WITH ASSOCIATED PROBABILITY VALUES[a]

The Esthetic SP seen as more:	The Political SP seen as more:
$P < .001$	
Sincere, Intellectual, Imaginative, Mature, Creative, Sympathetic, Good sense of values	Conformist, Sociable Forceful, Mercenary
$P < .01$	
Logical	Frivolous, Well-rounded, Shallow, Affected
$P < .05$	
Impractical	Pretentious, Authoritarian

[a] As determined by t tests ($df = 71$) of flattery discrepancy scores obtained by subtracting each *S*'s flattery score on the political SP from his flattery score on the esthetic SP and comparing the grand mean with the expected value of zero.

Differences between treatments. While not crucial to the main hypothesis involving individual differences, it was expected that both conditions of anticipated interaction would enhance the appeal of the esthetic SP relative to the political SP. This expectation was quite dramatically *not* confirmed by the results. The three treatments differ from each other with regard to flattery of the political SP and the discrepancy between political and esthetic SPs (both at $p < .05$ by analysis of variance), but the political person was more flattered and preferred when interaction was anticipated and the esthetic person was less preferred. As may be seen in Table 2, the AICoop condition was most divergent from the control condition, with the AIComp condition falling in the middle. The choice data show this same differential preference for the political SP in the interaction conditions as

TABLE 2
MEAN FLATTERY SCORE[a] OF EACH STIMULUS PERSON BY CONDITION AND TESTS OF DIFFERENCES BETWEEN MEANS ACROSS CONDITIONS

Condition	Stimulus Person		
	Political	Esthetic	Discrepancy (E−P)
Control			
$\bar{X}$	111.44	136.74	25.30
SD	6.04	4.19	27.35
AIComp			
$\bar{X}$	120.52	131.17	10.65
SD	7.43	5.54	27.04
AICoop			
$\bar{X}$	127.45	125.91	−1.54
SD	5.29	5.09	23.87
Source of Difference			
F: Between/Within	3.85 ($P < .05$)	2.91 ($P > .05$)	5.65 ($P < .01$)
t: Control vs. AICoop	3.06 ($P < .01$)	2.57 ($P < .05$)	3.34 ($P < .01$)
Control vs. AIComp	1.49 ($P > .05$)	1.27 ($P > .05$)	1.74 ($P > .05$)
AICoop vs. AIComp	— ($P > .05$)	1.05 ($P > .05$)	1.52 ($P > .05$)

[a] The larger the mean score, the more favorable the impression. A score of 120 (4 × 30) represents an ideally neutral impression.

compared with the control condition, but not as dramatically and not to a significant extent.

In an attempt to clarify further the differences between treatments, simple analyses of variance were run on each trait for each *S*'s discrepancy in rating the political and esthetic SPs to test the significance of the between-conditions effect. The difference between conditions was significant for 8 of the 30 traits. As compared with the control *S*s, the AICoop *S*s judged the political SP to be more Sincere, Logical, Sympathetic, Dedicated, and to have a better Sense of Values—all relative to the esthetic SP. The esthetic SP was seen as more Misguided, Pretentious, and Shallow relative to the political SP. With only one exception (where their mean equals that of the control group) *S*s in the AIComp group fall almost precisely in the middle. These findings offer strong support for using the flattery index as a major dependent variable because they suggest high intertrait reliability. They shed little further light on the reasons why the flattery scores differ as a function of experimental treatment, however.

Political values, Machiavellianism, and Flattery scores. The basic hypothesis of the present research predicted that the anticipation of interaction would alter the correlation between AVL and Mach IV scores on the one hand and flattery of the political SP on the other. Specifically, these correlations should be significantly lower or more negative when interaction was anticipated. Differential predictions were not made for AVL and Mach IV scores, but it was expected that the predicted decline in correlation would be more strikingly borne out for the Mach IV Scale which is a more direct measure of manipulative orientation than AVL. Nor were separate predictions made for the two types of anticipated interaction, although it was expected that simple choice of SP would be more highly correlated with flattery in the AICoop than the AIComp condition.

Product-moment correlations between AVL score and the two flattery scores, and between Mach IV score and the two flattery scores are presented by condition in the odd columns of Table 3. While none of the 12 correlations differ significantly from zero, the direction of difference is as predicted for the Mach IV Scale correlations with flattery of the political SP. The difference between the AIComp and control conditions approaches significance. Equally striking, however, is the fact that AVL score in the AIComp condition shows an almost significant positive correlation with flattery of the political SP. The AVL and the Mach IV correlations in this row differ significantly from each other ($p < .01$; Walker & Lev, 1953, p. 256–257). All the correlations between flattery of the esthetic SP and scores on the two political scales are negative but

TABLE 3
RELATIONSHIP BETWEEN FLATTERY SCORES, AVL-POLITICAL, AND MACH BY EXPERIMENTAL CONDITION
(Product-Moment and Partial Correlations)

Condition	*n*	Flattery of: Political SP				Esthetic SP			
		AVL[a]	AVL.M[b]	M[a]	M.AVL[b]	AVL[a]	AVL.M[b]	M[a]	M.AVL[b]
Control	27	.106	.144	−.042	−.122	−.260	−.322	−.087	.053
AIComp	23	.358	.576*	−.327	−.667**	−.274	−.229	−.190	−.113
AICoop	22	.121	.192	−.091	−.203	−.152	−.089	−.154	−.091

[a] Product-moment correlation coefficients.
[b] Partial correlation coefficients.
* Significant at the .01 level.
** Significant at the .001 level.

low, and weakly suggest that esthetic persons were more likely to be flattered by those low in political orientation.

Since the correlations reported in the first and third columns of Table 3 are opposite in sign, and since AVL score shows a moderate positive correlation with Mach IV score, it is apparent that all six of the correlations involving flattery of the political SP are underestimations of the relationship with one scale, uncontaminated by influence from the other. In other words, the fact that Mach IV and AVL are positively correlated makes it difficult for any of the correlations to depart too far from zero. The situation is one which calls for computation of partial correlations between flattery score and each scale score, holding constant the effects of the other scale score. The partial correlations are presented in the even columns of Table 3 for all 12 cells. The same trends are apparent, only now the two correlations for the political SP in the AIComp condition differ significantly from zero. Furthermore, the correlation involving the Mach IV Scale differs significantly from the Mach-flattery correlation in the control condition.

To *summarize* these results, the higher one's political values (AVL) the more he tends to flatter the political SP when he anticipates a competitive interaction with him. The higher one's tendency to be "Machiavellian," the *less* he tends to flatter the political SP if competitive interaction is anticipated. The relationship is magnified when the tendency to value political prestige (AVL) is held constant. The anticipation of a cooperative interaction results in the same level of correlation as that in the control condition.

Preference and Flattery

The preference data more or less parallel the flattery scores when examined by treatments alone, showing a nonsignificant tendency for less differential preference for the esthetic SP in the interaction conditions as compared with the control condition. Since *S*s were asked to indicate which of the SPs they *liked best* in the control condition, and which of the SPs they *wished to interact* with in the two experimental conditions, the results of the two kinds of preference data should be treated separately.

In the control condition, "liking" is highly correlated with a tendency to flatter the liked SP ($r_{pb} = .743$, $p < .001$). Looking at the individual differences data, the higher the political values the greater the tendency to "like" the political SP ($r_{pb} = .426$, $p < .05$). The correlation between Mach IV Scale score and "liking" is very low and not significant. Restricting the interpretation to the AVL measure of political values, then, the predicted similarity hypothesis is borne out more strongly with the sociometric preference than with the flattery data in the control condition.

In the interaction conditions, "choice" also relates to some extent to flattery score. A secondary hypothesis predicted that the correlation between choice and flattery would be higher in the AICoop than in the AIComp condition. This prediction is based on the assumption that one might prefer to interact

with a disliked person on a competitive task, whereas the generalization from choice to positive general evaluation would be greater when cooperation is anticipated. The point biserial correlations between flattery discrepancy score (see Footnote a, Table 1) and preference were .670 for the AICoop condition and .378 for the AIComp condition. While these correlations do not differ significantly from each other, the former is significantly different from zero ($p < .001$) while the latter is not ($p > .05$).

None of the correlations relating choice of partner to AVL-political score or Mach IV Scale score is significant in either interaction condition.

DISCUSSION

The results of the present study raise a number of problems of interpretation. While the experiment was generally successful in demonstrating the effects of each independent variable, the direction of these effects was not always as predicted.

The first finding of note was that the anticipation of interaction resulted in more positive evaluation of the political SP and more negative evaluation of the esthetic SP. This effect was particularly evident in the condition where a cooperative interaction was anticipated. The most plausible explanation for this reversal of expectation centers around perceptions of the SP's sociability and level of apparent motivation. In general, across all conditions, more positive characteristics were attributed to the esthetic than to the political SP. Apparently, the typical *S* saw the political SP as rather shallow and one-dimensional, but as highly sociable and gregarious. The esthetic SP was seen as aloof and distant in spite of his greater depth and breadth. This difference in sociability seems to be evident in the impressionistic notes made by the *S*s immediately after hearing the interviews, though these notes were not formally coded. When interaction is anticipated, the ease of conversing with the SP presumably depends on his sociability which therefore takes on a special significance. The political SP, we may thus infer, is seen as easier to interact with in the short run, and this short-run expectation generates a more favorable halo effect in rating the SP.

There are two reasons why the political SP may be especially preferred in the AICoop condition. One emphasizes the enhanced value of sociability in a cooperative situation. A second explanation concerns the impression of greater drive and tenacity in the political SP, again as suggested by the *S*s' notes. In spite of his presumably greater knowledge of "literature, music, and the arts," the esthetic SP is more uncertain of his goals and less persevering in their attainment. It might be added that several *S*s directly questioned the esthetic SP's breadth of knowledge, conveying their impression of him as a superficial pseudo-intellectual. There is some reason for believing, then, that the political SP is seen as a better team member because of his tenacity and seriousness of purpose, and that the superior knowledge of the esthetic SP is not great enough to offset this motivational edge.

The most important findings of the present experiment concern the role of individual differences in the prediction of flattery and interpersonal choice. The expectation that prompted the experiment was that political people like other political people (and flatter them) in the abstract but not in the flesh, or when a relevant interaction is anticipated. The experimental results show that it is most important to clarify further the meaning of "political." If the referent is political *values*, there is generally support for the similarity hypothesis. That is, the person who values political things in general, also values people who have a manifestly political orientation. When interaction was not anticipated, the higher a person's political values, the greater the tendency to "like" the political SP more than the esthetic SP. In this control condition, however, there was no corresponding tendency toward differential flattery of the political SP. When a competitive interaction was anticipated, the positive attributes of the political SP became more salient to *S*s with high political values, but while the "highs" flattered him more, they did not show any differential tendency to choose him in the interaction. Neither flattery nor preference related to political value score when a cooperative interaction was anticipated.

These results are not easy to rationalize in

detail, and it is particularly difficult to see why the more political *S*s flattered the political SP more when a competitive than when a cooperative interaction was anticipated. This finding may in some way reflect the fact that the content of the discussion in AIComp (leadership) is of greater interest to the high political *S* than the content in AICoop (literature, music, etc.).

The results for the Mach IV Scale are somewhat more coherent and generally support the importance of complementarity when, as predicted by the major hypothesis, interaction is anticipated. In other words, people with a strong need to manipulate their social environment tend to devalue others with similar orientation when a competitive interaction is anticipated. While the relationship is not significant for the AICoop condition, there is good reason to suppose that the negative relationship between Mach IV Scale score and flattery is not restricted to competitive interaction settings. A glance at the scatter diagram relating Mach IV to flattery in the AICoop condition reveals two very atypical *S*s, removal of which results in a significant correlation of $-.471$. Even with these two *S*s, an analysis of variance relating three levels of Mach IV score to flattery is significant at the .01 level, with high and middle thirds of Mach IV showing strikingly different means in the predicted direction ($p < .001$ by t test). With some reservations concerning the low Mach IV *S*s in the AICoop condition, then, the anticipation of interaction in general tends to decrease the *S*s' evaluation of the SP most similar to the self.

Perhaps the most striking feature of the results is that political values, as measured by the AVL, and Machiavellianism predict to flattery score in the opposite direction in spite of the fact that they are positively correlated with each other. Since it was the Mach IV Scale, rather than the AVL, which confirmed the major experimental prediction, this dimension surely seems worthy of further investigation.

SUMMARY

In an effort to shed further light on the conditions favoring similarity and complementarity in social preference, an experiment was designed to investigate the perception of a political stimulus person (as compared with an esthetic stimulus person) by *S*s varying in the strength of their own political orientation. This "orientation" was measured both by the political scale of the Allport-Vernon-Lindzey Study of Values, and by Christie's Mach IV Scale. In two conditions of the experiment, *S*s were led to anticipate interacting with one of the stimulus persons, while in a control condition the stimulus was pointedly described as tape-recorded so that there would be no expectation of ever meeting the SP.

As predicted, the higher the *S*'s score on the Mach IV Scale, the more negative the evaluation of the political stimulus person *when* interaction was anticipated. However, the higher the *S*'s score on the political scale of the AVL, the more positive the evaluation of the political stimulus person, especially when a competitive interaction was anticipated. Thus, whether the similarity or complementarity hypothesis is valid depends not only on the particular personality characteristic being considered but, also, on the interaction context of evaluation.

Contrary to the prediction, the political stimulus person was evaluated more favorably and the esthetic stimulus person less favorably when interaction was anticipated. This result appears to be a function of the greater sociability of the political stimulus person. A number of other findings, more or less unexpected, were presented and discussed.

REFERENCES

ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.

ALLPORT, G. W., VERNON, P. E., & LINDZEY, G. *Study of values*. (rev. ed.) Cambridge: Houghton Mifflin, 1951.

BOWERMAN, C. E., & DAY, BARBARA R. A test of the theory of complementary needs as applied to couples during courtship. *Amer. social. Rev.*, 1956, **21**, 602–605.

CHRISTIE, R., & MERTON, R. K. Procedures for the sociological study of the values climate of medical schools. *J. med. Educ.*, 1958, **33**, 125–153.

JONES, E. E. Authoritarianism as a determinant of first-impression formation. *J. Pers.*, 1954, **23**, 107–127.

JONES, E. E., & THIBAUT, J. W. Interaction goals as bases of inference in interpersonal perception. In R. Taguiri and L. Petrullo (Eds.), *Person percep-*

tion and interpersonal behavior. Stanford: Stanford Univer. Press, 1958. Pp. 151–179.

LINDZEY, G., & BORGATTA, E. F. Sociometric measurement. In G. Lindzey (Ed.), *Handbook of social psychology*. Cambridge: Addison-Wesley, 1954. Pp. 405–448.

PRECKER, J. A. Similarity of valuings as a factor in selection of peers and near-authority figures. *J. abnorm. soc. Psychol.*, 1952, **47**, 406–414.

ROSOW, I. Issues in the concept of need-complementarity. *Sociometry*, 1957, **20**, 216–233.

VAN DYNE, VIRGINIA E. Personality traits and friendship formation in adolescent girls. *J. soc. Psychol.*, 1940, **12**, 291–303.

WALKER, HELEN M., & LEV, J. *Statistical inference*. New York: Holt, 1953.

WINCH, R. F. *The modern family*. New York: Holt, 1952.

WINCH, R. F. The theory of complementary needs in mate selection: A test of one kind of complementariness. *Amer. sociol. Rev.*, 1955, **20**, 52–56.

WINCH, R. F., KTSANES, T., & KTSANES, VIRGINIA. The theory of complementary needs in mate selection: An analytic and descriptive study. *Amer. sociol. Rev.*, 1954, **19**, 241–249.

ZIMMER, H. Motivational factors in dyadic interaction. *J. Pers.*, 1956, **24**, 251–262.

(Received July 28, 1958)

MASCULINITY, IDENTIFICATION, AND FATHER-SON RELATIONSHIPS

PAUL MUSSEN AND LUTHER DISTLER
University of California, Berkeley

BY IDENTIFYING with his parents—i.e., endeavoring "to mold (his) own ego after the fashion of one that has been taken as a model" (S. Freud, 1949, p. 63)—the child begins to acquire his parents' personality characteristics, behavior, values, motivations, and attitudes. Two of the major consequents of the processes of identification are the development of the superego, or conscience, and the acquisition of behavior and personality characteristics appropriate to his own sex (sex-typing).

It has been hypothesized that both boys and girls form their first identification with a female, the mother, because she is likely to be the most gratifying and affectionate person with whom the child has contact (Mowrer, 1950; Sears, Maccoby, & Levin, 1957). Continued identification with the mother is appropriate for a girl, for she must acquire feminine personality characteristics and interests, but the boy "must shift to a masculine identification, sometime in his early years, if he is to develop normally masculine personality" (Sears et al., 1957, p. 373).

According to psychoanalytic theory, the boy's shift to identification with his father begins during the Oedipal phase of development and is motivated by fears and anxieties related to his hostility toward that parent. By identifying with his father, the boy's fears of counter-aggression are reduced and, at the same time, he vicariously obtains his mother's attention and affection. This is the defense mechanism that has been called "identification with the aggressor" (A. Freud, 1937) and "defensive identification" (Mowrer, 1950).

In contrast to this, the developmental hypothesis states that identification with the father depends on a positive, affectionate relationship between father and son (Mowrer, 1950; Payne & Mussen, 1956). If the father is an important source of nurturance, reward, and satisfaction, his responses and characteristics acquire secondary reward value and the boy imitates his father's behavior in order to "reproduce bits of the beloved and longed for parent" (Mowrer, 1950, p. 615).

Although the two hypothesized explanations of the dynamics underlying identification with the father seem to be vastly different from each other, there is some supportive evidence for each of them. The validity of the developmental identification hypothesis has been tested in a number of recent studies (Levin & Sears, 1956; Payne & Mussen, 1956; P. Sears, 1953; Sears et al., 1957; Sears, Pintler, & Sears, 1946). Thus in one study, five-year-old boys with warm and affectionate fathers (as judged by home interviews) showed stronger father identifications than did boys with "cold" fathers (P. Sears, 1953). It has also been reported that adolescent boys who perceive their fathers as nurturant (according to projective tests) are more likely to identify strongly with this parent than are boys who do not perceive their fathers as rewarding (Payne & Mussen, 1956). Moreover, certain characteristics of boys generally regarded as consequents of identification with the father—sex-typed behavior such as aggression (Levin & Sears, 1956), masculine interests and attitudes (Payne & Mussen, 1956), and highly developed conscience (Sears et al., 1957)—tend to be associated with favorable father-son relationships.

The hypothesis of defensive identification is a complex one that is difficult to investigate empirically. However, there are extensive clinical observations indicating that, as this hypothesis would predict, many children attempt to reduce their anxieties by adopting the aggressive characteristics of individuals whom they perceive as threatening to them (A. Freud, 1937; Mowrer, 1950).

It is quite possible, of course, that the two hypothesized identification processes are not mutually exclusive but may function together to facilitate the boy's shift from a feminine to a masculine identification (Sears et al., 1957). This view is consistent with that of the role theorists who equate identification with the

father with "father role playing" (Brim, 1958; Cottrell, 1942; Parsons, 1955). These theorists maintain that identification, or role-playing, depends on the *power* of the identificand—a combination of his reward value *and* his threat or punishment potential. More specifically:

> Given two . . . persons with whom one interacts and who differ in power over the actor (the identifier), i.e., differ in the degree to which they control rewards and punishments for the actor, one would predict that the actor would adopt many of the characteristics of the powerful as contrasted to the less powerful other person. This follows from the fact that it is more important to the actor to predict the behavior of the powerful figure, that he is motivated more strongly to take his role (i.e., to identify), that rewards and punishments are more impressive and the learning consequently better (Brim, 1958, p. 3).

The present paper reports the results of an investigation that attempted to evaluate the validity of the three hypothesized explanations of identification—developmental, defensive, and role-playing. It was assumed that, for young boys, appropriate sex-typing of interests is an indication of identification with the father, and that amount of identification can be estimated from the degree of sex-typing. The rewarding-nurturant and threatening-punitive qualities of the father were evaluated in terms of the child's *perceptions* of the father, rather than in terms of the father's actual behavior, the criterion used in some previous studies (Levin & Sears, 1956; P. Sears, 1953; Sears et al., 1957).

There were several central questions. How do boys who are strongly identified with their fathers, i.e., highly masculine in their interests, perceive their fathers: as basically nurturant and rewarding, as would be predicted from the developmental identification hypothesis, or as punitive and threatening, in accordance with the defensive identification hypothesis? Or are they viewed as powerful agents of both reward and punishment, as the role-playing hypothesis of identification maintains?

METHOD

Subjects. According to psychoanalytic theory and observational data, the late preschool period is a critical time in the boy's shift from feminine to masculine identification. For this reason, it seemed best to use kindergarten boys as *Ss* for this investigation.

Initially, 38 white five-year-old boys in two kindergarten classes of a predominantly middle-class public school[1] were given the IT Scale for Children (ITSC), a projective test of sex-role preference (Brown, 1956). The test consists of 36 picture cards depicting various objects, figures, and activities commonly associated with masculine or feminine roles. The child is given a card with a figure drawing unstructured as to sex and referred to as IT. He is then presented with groups of pictures of toys and with paired choices of activities, and asked to choose what IT would like.

"In using IT the assumption is made that the child will project himself or herself into the IT-figure on the basis of his own or her own sex-role preference, and will attribute to IT the child's own role preference." (Brown, 1956, p. 5). The possible range of scores on the test is from zero, an exclusively feminine score, to 84, an exclusively masculine score.

The 10 boys with the highest scores (range 79–84, with a mean of 83) and the 10 boys with the lowest scores (range 30–63, with a mean of 53) were selected for further study.[2] The two groups were matched in socioeconomic status, each of them having the same number of upper-lower, lower-middle, and upper-middle class boys. Ten of the *Ss*, 5 high scorers and 5 low scorers, came from one kindergarten class, and 10 from the other.

Measures of parent-child relations. Between one and four weeks after he had taken the ITSC, each of the 20 boys was tested individually in a structured doll-play session. Early in the session, the child familiarized

---

[1] The authors wish to express their appreciation to A. B. Campbell, Assistant Superintendent of Schools of Berkeley, C. B. Johnson, Principal of the Jefferson School, and C. B. Holmes and E. P. Light, kindergarten teachers, for their cooperation in this study.

[2] While the *Ss* were selected only on the basis of masculinity score on the ITSC, there is some suggestive evidence that those scoring low on the test were also somewhat more feminine in personality characteristics. Their teachers rated the *Ss* on 20 scales of personality characteristics, adapted, with slight modifications, from the Fels Child Behavior Scales. These characteristics were: aggression, affection, cheerfulness, competition, conscience, curiosity, emotional control, gross activity, friendliness, gregariousness, kindness, leadership, obedience, patience, physical apprehensiveness, quarrelsomeness, sensitivity, shyness, suggestibility, and tenacity (Richards & Simons, 1941).

The two groups of *Ss* differed significantly (at approximately the .02 level) on 4 of the 20 scales. The low scorers rated significantly higher than the highly masculine on: gross activity, friendliness, cheerfulness, and gregariousness. The latter three characteristics are, according to sociological analyses of sex role, specifically assigned to the expressive or feminine role (Brim, 1958; Cottrell, 1942; Parsons, 1955). The relatively high ratings of gross activity among those low in masculinity may also reflect group differences in these social characteristics. Significant positive correlations between the gross activity scale and the cheerfulness and gregariousness scales have been reported previously (Richards & Simons, 1941). Apparently, the child who is expressive, outgoing, and cheerful is quite likely to be seen as more active.

himself with three easily manipulated dolls, representing a mother, a father, and a boy doll, and some very simple toy furniture, which were on a low table in the testing room. *S* was told that these were to be used in a story-telling game in which the examiner would make up the first part of the story and *S* would complete it.

The examiner then presented, with appropriate manipulations of the dolls, a series of nine incomplete, family situation stories, and the child was asked to complete each one in play. The stories were structured in such a way that the child could depict either or both parents as nurturant and/or punitive. Five of the incomplete stories follow:

1. The child wants a certain toy. He can't reach it. He goes into the living room to get help. Both Mommy and Daddy are busy reading. What happens?

3. The child lives on a very busy street. Mommy and Daddy told him never to cross the street alone. The child is playing in the front yard, and Mommy and Daddy are not there. A friend is on the other side of the street playing with his new bike. The child wants to cross the street very much. What happens?

4. The child is having fun playing with his toys. Mommy and Daddy say "It's time to go to bed now." The child says "I don't want to go to bed now." Then the child throws a toy on the floor and it breaks. What happens?

5. The child is getting ready to go to school. He has a knot in his shoelace. He can't fix it. What happens?

8. Let's pretend the little boy had a bad dream. Now the little boy wakes up from his bad dream, screaming. He calls for Mommy or Daddy. Which one does he call for, Mommy or Daddy? Then what happens?

If the child failed to respond to one of the stories or said "I don't know," the story was repeated or the question rephrased. If *S* did not mention the parents in his play, the examiner asked "What did Mommy or Daddy say or do?" When the child did not designate a specific parent in his response, the examiner inquired "Who did that, the Mommy or the Daddy?" Each doll-play session was completely recorded.

Scoring the doll-play responses. The major hypotheses with which the study was concerned dealt with the relationship between boys' masculinity scores and their perceptions of their fathers. The structured doll-play situation, however, permitted evaluations of the child's perceptions not only of his father but also of his mother and his parents as a unit (when mentioned without specific designation of mother or father) as nurturant and/or punishing. The assumption underlying the use of this technique is that the boy's responses in doll play reveal his own feelings about his parents' treatment of him.

Each story was scored for the presence of nurturance or punishment by the father, mother, or "they" (parents undifferentiated). The stories were scored without the scorer's knowledge of the child's ITSC score.

The following scoring categories were used:

Father Nurturance (FN) score was the total number of stories in which the child character in the stories received help, comfort, attention, or reassurance from the father (e.g., "his Daddy gets it for him" in response to Story 1 described above).

A Mother Nurturance (MN) score was derived by counting the number of stories in which the mother was depicted as nurturant. (For example, in response to the story about the dream, "He calls his Mommy and she says it's just a dream and he can sleep with her.")

They Nurturance (TN) score was the number of stories in which the parents as a unit were nurturant. (For example, *S*, in response to Story 3: "They help him cross the street." *E:* "Who helps, Mommy or Daddy?" *S:* "His Mommy and Daddy.")

Total Nurturance (TotN) score was the sum of the MN (Mother Nurturance), FN (Father Nurturance), and TN (They Nurturance) scores.

The total number of stories in which the father, mother, or "they" disciplined, spanked, criticized, or admonished the child in the story constituted, repectively, the Father Punishment (FP), Mother Punishment (MP), and They Punishment (TP) scores.

Total Punishment (TotP) score was the sum of the FP (Father Punishment), MP (Mother Punishment), and TP (They Punishment) scores.

It should be noted that a given story could be scored for more than one variable. For example, if the mother in the story spanked the child and later comforted him, the story would be scored both Mother Punishment (MP) and Mother Nurturance (MN).

The total number of stories involving relationships with the father, either nurturant or punitive (i.e., FN plus FP) constituted the Father Power (FPow) score. Analogously, Mother Power (MPow) was the sum of the Mother Nurturance (MN) and Mother Punishment (MP) scores.

RESULTS

The three theoretical formulations of the dynamics underlying the shift from feminine to masculine identification, outlined above, could be evaluated by determining the relationships between masculinity status and perceptions of parents, as measured by the child's responses in doll play. If the developmental hypothesis is valid, highly masculine boys would perceive their fathers as more rewarding and affectionate or, operationally, would have higher Father Nurturance (FN) scores than boys with low masculinity scores. On the basis of the defensive identification hypothesis, however, it would be predicted that boys who were most strongly identified with their fathers —as this is reflected in their highly masculine interests—would regard their fathers as more threatening and punitive, i.e., would have higher Father Punishment (FP) scores than boys whose father identifications are weak. The role-playing hypothesis of identification offers a third prediction. According to this

theory, boys with strong tendencies to play the father role or to identify with the father would feel that their fathers were powerful figures and important sources of both rewards and punishments. In terms of the variables of this study, this would mean that highly masculine boys would have higher Father Power (FPow) scores than boys who have not achieved strong masculine identifications.

Since responses to the structured doll-play situation were scored for perceptions of the mother and "they" (parents as a unit) as nurturant-gratifying and/or punitive-threatening, it was also possible to assess the relationships between these variables and strong or weak sex-typing among five-year-old boys.

Table 1 presents the mean scores of the high and low masculinity groups on all the doll-play scores.

In view of the facts that the number of *Ss* in each group was small and that the distributions of scores on these variables was nonnormal, *U* tests (Mann & Whitney, 1947) were employed to compare rank transformation scores on all doll-play scores of *Ss* scoring high and low on the ITSC. The results of these tests and their significance levels are summarized in Table 2.

It is obvious from this table that the two groups of *Ss* differ significantly in many of their perceptions of their families. Compared with boys low in masculine identification, as measured by the ITSC, those who were high in this characteristic perceived themselves as receiving more Total Nurturance (TotN) (i.e., more nurturance from all sources combined). Evidently, this difference is primarily attributable to differences in one major component of TotN, the perceptions of father's nurturance, for the two groups differ significantly in Father Nurturance (FN) scores but not in the Mother Nurturance (MN) or They Nurturance (TN) scores.[3]

This finding seems clearly consistent with the developmental identification hypothesis. It supports the prediction, made on the basis of this hypothesis, that young boys are more likely to identify strongly with their fathers, and thus to acquire masculine interests, if they perceive their fathers as highly nurturant and rewarding.

The data also appear to lend support to the second, or defensive identification, hypothesis, which predicted that high masculine identification would be related to views of the father as threatening and punitive. Highly masculine boys tended to attribute more punishment to the fathers in their doll-play stories than boys low in masculine identification did, although the difference between the two groups in the father punishment score was not quite as marked as the difference in the father nurturance variable.

The third, or role-taking, hypothesis states that the degree of identification and, conse-

TABLE 1

MEAN SCORES OF BOYS HIGH AND LOW MASCULINITY GROUPS ON FAMILY PERCEPTION VARIABLES

Variable	High Masculinity Group	Low Masculinity Group
Mother Nurturance (MN)	2.1	1.7
Father Nurturance (FN)	3.7	2.2
They Nurturance (TN)	.8	.7
Total Nurturance (TotN)	6.6	4.7
Mother Punishment (MP)	1.2	1.5
Father Punishment (FP)	2.8	2.1
They Punishment (TP)	.2	.7
Total Punishment (TotP)	4.2	4.3
Mother Power (MPow)	3.3	3.2
Father Power (FPow)	6.5	4.3

TABLE 2

DIFFERENCES BETWEEN HIGH AND LOW MASCULINITY GROUPS ON FAMILY PERCEPTION VARIABLES

Variable	*U*	*p*	Group with Higher Scores
Mother Nurturance (MN)	45.5	NS	—
Father Nurturance (FN)	23.5	.02	Highs
They Nurturance (TN)	48.5	NS	—
Total Nurturance (TotN)	16.0	.004	Highs
Mother Punishment (MP)	44.0	NS	—
Father Punishment (FP)	30.5	.06	Highs
They Punishment (TP)	34.0	.07	Lows
Total Punishment (TotP)	47.0	NS	—
Mother Power (MPow)	49.5	NS	—
Father Power (FPow)	18.5	.007	Highs

[3] While the use of *t* tests was not entirely warranted because of the nonnormality of the doll-play scores, these tests were applied to the data of Table 1. The results were very similar to the *U* test results, i.e., the high and low masculinity groups differed significantly from each other in the variables FN ($p = .01 - .025$), TotN ($p = <.01$), FP ($p = .01 - .025$), FPow ($p = <.005$) and almost significantly in TP ($p = .05 - .10$).

quently, sex-role learning varies with the amount of the child's interaction with the identificand and the degree to which the latter has power over him, i.e., controls both his rewards and his punishments. The data of the present study seem to be fully in accord with this hypothesis, since the low and high masculinity groups differ markedly in Father Power (FPow) scores, the high identifiers giving a significantly greater number of these responses. This finding is exactly what would be anticipated, since this score is composed of the FN and FP scores, and the highly identified group scored significantly higher in each of these.

The high and low masculinity groups were not significantly different in any of the variables related to perceptions of the mother. However, compared with those scoring low on the ITSC, high scorers tended to perceive their parents as a unit as less punitive, i.e., tended to have lower They Punishment (TP) scores ($p = .07$). This would seem to be further confirmation of the findings of another study which concluded that adolescent "boys who feel comfortable in their relationships with their parents adopt more of their father's behavior and attitudes than boys who experience less favorable parent-child relationships" (Payne & Mussen, 1956, p. 361).

It may be that the boy who scores high in They Punishment (TP) views his family milieu as hostile and unfriendly. While he may not feel that either parent is particularly threatening, he may feel generally rejected, unwanted, and unimportant. Insofar as this is true, the child may be expected to attempt to avoid intensive interactions with his parents. Under these circumstances, he should not identify strongly with his father and, consequently, should not acquire highly masculine interests.

DISCUSSION

The data of the present study indicate that for boys, sex-typing of interests is more directly related to their perceptions of their fathers than to perceptions of their mothers. This finding is in accord with the findings of previous studies (Levin & Sears, 1956; Payne & Mussen, 1956; Sears et al., 1957) as well as with an assumption underlying all three of the identification hypotheses outlined above; namely, that the acquisition of masculine interests, attitudes, and patterns of behavior is primarily determined by the boy's interactions with his father.

Some of the findings of this study lend support to the defensive identification hypothesis; others seem to support the developmental hypothesis. Since the two hypotheses generate predictions that are in some respects diametrically opposed, there are also data that are inconsistent with both hypotheses. Thus, since the developmental hypothesis postulates that identification with the father is dependent upon nurturant rewarding interactions with that parent, the finding that boys also identify with punitive, threatening fathers must be inconsistent with this hypothesis. Conversely, the finding that boys are more likely to identify with a nurturant father is inconsistent with the view of defensive identification or identification with the aggressor.

A high level of masculine identification does not appear to depend on any one specific type of father-son relationship. From the child's point of view, the significant factor seems to be the father's *salience*—his importance in the child's life—rather than the particular techniques he uses in dealing with his child. Thus, as a group, boys who have made substantial father-identifications—reflected in their strongly sex-typed interests—perceived their fathers as both more nurturant *and* more punitive. Hence, masculine sex-role identification cannot be attributed exclusively to either the reward value or the threat potential of the father.

For these reasons, it seems to us that the role theory of identification is most fully consistent with—and most adequately integrates—the present data. The two groups were clearly differentiated on the Father Power (FPow) score, indicating that, compared with the other group, the highly masculine boys had—or at least perceived that they had—more intensive interactions with their fathers. This is exactly what is predicted by the role theory of sex-role identification, which maintains that more interaction with another individual, e.g., the father, leads to greater assimilation of his role.

Moreover, role theory states most explicitly that an individual is most likely to assimilate the role of, or identify with, individuals he sees as powerful. From the child's point of view, the most powerful individual is probably the one who most effectively controls his rewards and punishments. In this way, the theory implies that both reward and punishment strongly influence the course of role learning.

It follows from these postulates of role theory that the boy should be most strongly motivated to imitate or to practice his father's role frequently if he has a great deal of interaction with his father and sees him as a powerful source of rewards and punishments. Under these circumstances, the child gets extensive experience playing the father's role and adopts more of the father's characteristics, including those connected specifically with his sex-role. Our data are substantially in accordance with these theoretical predictions. The highly father-identified *Ss*—those who had adopted masculine interests and behavior—did in fact perceive their fathers as more interactive with them and as major providers of rewards *and* punishments.

There is also a psychoanalytic hypothesis which these data seem to confirm. In summarizing psychoanalytic writings on identification, Bronfenbrenner (1958) notes that one of the "syndromes of parent-child relationship . . . predisposing the child to incorporate or introject the parent" is "a relationship based on conditional love, in which the parent, willfully or unconsciously, withholds expression of affection as the price of conformity" (p. 128). This means that the boy is most likely to identify with his father if his feelings toward him are affectionate, while the father's love is given conditionally. According to this hypothesis, high nurturance from the father, together with the threat of withdrawal of his love (high FP) would lead to strong father identification. This prediction seems to be verified by the finding that highly masculine boys have relatively high Father Power scores, indicating a high degree of combined father nurturance and father punishment.

The findings may also be conceptualized in terms of general behavior theory. Among those with whom the preschool boy has intimate associations, the father is the one who has the most adequate knowledge of appropriate masculine behavior. If there is a high level of father-son interaction (high FPow scores), the father should frequently, and fairly regularly, reward the son's sex-appropriate responses and punish sex-inappropriate responses when they occur. Consequently, the boy's masculine responses should be relatively rapidly and effectively strengthened, while his sex-inappropriate responses become extinguished. In short, vigorous application of both rewards and punishments by the father facilitate the son's shift from feminine to masculine identification.

On the other hand, the condition of relatively little father-son interaction implies sporadic, and at best ineffective, rewards and punishments by the father. Under these circumstances, the shift in identification may be a difficult one, and the child's acquisition of sex-typed interests and behavior may be considerably retarded.

It seems that the father's use of *both* reward *and* punishment is his most effective method of "teaching" his son masculine behavior. It must be emphasized, however, that the high masculine and low masculine groups also differed significantly in the individual variables relating to perceptions of father reward and father punishment. From this it may be inferred that sex-appropriate behavior may be 'taught" primarily by rewarding appropriate responses *or* punishing inappropriate responses. In short, from the child's point of view, the important factor is the salience of his relationship with his father, not the particular techniques that the father uses in handling him.

Of course, the present data refer *only* to masculinity of interests. Other possible long-range consequents of different processes of identification cannot be evaluated. For example, it is quite possible that boys who learn their sex roles primarily as a result of being punished or threatened by their fathers differ in many ways (e.g., in personality characteristics, self-conceptions, or basic motivations) from those who achieve their masculine identifications by means of developmental identification or through learning based on some combination of reward and punishment by the father. As the data stand, however, we can

only state that five-year-old boys may shift from feminine to masculine identification, as measured by the acquisition of masculine interests, as a result of either developmental or defensive identification processes or, as role theory suggests, by some combination of the two.

SUMMARY

The IT Scale for Children (ITSC), a test of sex-typing of interests, was administered to 38 white, middle-class, kindergarten boys. The 10 *Ss* scoring highest in the test were assumed to have developed the highest degree of male role identification, while the 10 scoring lowest were considered the least strongly identified with this role.

In order to determine the relationship between parental perceptions and degree of masculine identification, each of the 20 boys was tested in a structured doll-play situation. During the session, the child completed, in play, nine incomplete stories involving parent-child relations. Responses were scored in terms of the amount of nurturance, punishment, and power (nurturance plus punishment) attributed to the mother, father, and parents as a unit.

The study was designed to evaluate three hypothesized processes of identification—developmental, defensive, and role-taking. Analysis of the data provided evidence consistent with all three hypotheses. As predicted from the developmental identification hypothesis, young boys who were strongly identified with the male role perceived their fathers as more rewarding and nurturant ($p = .02$) than their weakly identified peers did. According to the defensive identification hypothesis, the strongly father-identified boys perceive their fathers as more punitive and threatening. This hypothesis was also supported ($p = .06$). Boys high and low in masculinity were also clearly differentiated on the Father Power score ($p = .007$). This finding indicates that those who have made substantial male identifications view their fathers as powerful sources of *both* reward and punishment, and is in accordance with role theory, which maintains that the child is most likely to assimilate the role of an individual with whom he has intensive interactions, especially if this individual is powerful. To the present authors, it seems that role theory, with its explicit emphasis on the importance of both reward and punishment in role-learning, best integrates all these data.

REFERENCES

BRIM, O. G., JR. Family structure and sex role learning by children: a further analysis of Helen Koch's data. *Sociometry*, 1958, **21**, 1–16.

BRONFENBRENNER, U. The study of identification through interpersonal perception. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior*. Stanford, Calif.: Stanford Univer. Press, 1958. Pp. 110–130.

BROWN, D. D. Sex role preference in young children. *Psychol. Monogr.*, 1956, **70** (14, Whole No. 421).

COTTRELL, L. S., JR. The analysis of situational fields in social psychology. *Amer. sociol. Rev.*, 1942, **7**, 370–383.

FREUD, ANNA. *The ego and the mechanisms of defense*. London: Hogarth, 1937.

FREUD, S. *Group psychology and the analysis of the ego*. London: Hogarth, 1949.

LEVIN, H., & SEARS, R. R. Identification with parents as a determinant of doll play aggression. *Child Develpm.*, 1956, **27**, 135–153.

MANN, H. B., & WHITNEY, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. math. Statist.*, 1947, **18**, 50–60.

MOWRER, O. H. *Learning theory and personality dynamics*. New York: Ronald, 1950.

PARSONS, T. Family structure and the socialization of the child. In T. Parsons & R. F. Bales (Eds.) *Family, socialization, and interaction process*. Glencoe, Ill.: Free Press, 1955.

PAYNE, D. E., & MUSSEN, P. H. Parent-child relations and father identification among adolescent boys. *J. abnorm. soc. Psychol.*, 1956, **52**, 358–362.

RICHARDS, T. W., & SIMONS, MARJORIE. Fels Child Behavior Scale. *Genet. psychol. Monogr.*, 1941, **24**, 259–309.

SEARS, PAULINE. Child-rearing factors related to playing of sex-typed roles. *Amer. Psychologist*, 1953, **8**, 431. (Abstract)

SEARS, R. R., MACCOBY, ELEANOR E., & LEVIN, H. *Patterns of child rearing*. Evanston, Ill.: Row, Peterson, 1957.

SEARS, R. R., PINTLER, M. H., & SEARS, PAULINE. Effect of father separation on preschool children's doll play aggression. *Child Develpm.*, 1946, **17**, 219–243.

(Received August 6, 1958)

MINORITY GROUP BELONGING, SOCIAL PREFERENCE, AND THE MARGINAL PERSONALITY

D. W. LEWIT
University of Massachusetts

SINCE Lewin (1945) published his ideas on "self-hatred" in ethnically marginal persons, a number of relevant studies have appeared which have supported most of his contentions. Briefly, Lewin asserted that marginal men are frustrated by anti-minority prejudice in their continued efforts to become accepted by the privileged (dominant) minority or the majority, and remain uncertain of their group belonging. One consequence is a continuing state of high tension which in Lewin's system implies a generalized tendency to aggression, development of a negative attitude toward the perceived source of frustration, and/or de-differentiation of relevant aspects of the perceived environment. In the case of ethnically marginal persons, "self-hatred" refers to hostility toward the minority group as the perceived source of frustration, displaced from anti-minority members of the dominant group. The personal aspect of "self-hatred" is the assumed "unhappiness" (Lewin, 1945, p. 164) or personal insecurity theoretically associated with chronic marginal status.

The assumption that such marginal persons remain uncertain of their group belonging is supported by data from Steckler (1957) and Radke-Yarrow and Lande (1953). Steckler reports a positive correlation between anti-white and anti-Negro attitude scores in a sample of Negro students. Radke-Yarrow and Lande found a positive correlation between anti-Semitism and Jewish ethnocentrism scores in a Jewish sample. Glaser (1958), in his review of literature on ethnic identification, states that dominant group bigots and minority group chauvinists prefer to associate with members of their own group, while marginal persons are inconsistent in their social preferences. Sarnoff (1951) tested the hypothesis that anti-Semitic Jews, relative to non-anti-Semitic Jews, would show more personal characteristics similar to those of authoritarian "aggressors" with whom they presumably identified. Instead of showing uncritical glorification of parents, however, they tended (more than non-anti-Semitic *S*s) to fear and have negative attitudes toward parents. Also, rather than direct retaliation, the anti-Semitic Jews tended to show a relatively passive or self-hostile reaction to aggression directed against them from out-group members. Projective test data from the Radke-Yarrow and Lande sample were consistent with Sarnoff's findings. In general, these results are consistent with Lewin's assumptions.

It remains to be shown, however, whether marginal members of ethnic minorities necessarily or usually orient themselves positively toward the dominant or most privileged ethnic group. If an ethnic group less threatening and more accommodating than the dominant one is also available for social intercourse, it is reasonable to suppose that marginal members of an ethnic minority may orient themselves preferentially toward it. One purpose of the present study is to check this assumption. Another purpose is to determine whether marginal status necessarily implies personal insecurity, personal indiscriminateness (low cognitive differentiation), or socially inconsistent standards of perceptual judgment.

To these ends a social preference technique involving the expression of preferences for representatives of various ethnic groups in a familiar transient public situation was employed to measure orientation (operationally referred to hereafter as "social preference"), intra-group agreement, and intra-individual consistency. The Knutson (1952) Personal Security Inventory was used to provide a measure of insecurity close in meaning to Lewin's "unhappiness." Marginal Jews, non-marginal Jews, and Protestant North Europeans were used as *S*s to make preferential choices among North Europeans, Mediterraneans, Jews, and Negroes. The following hypotheses, stated so as to be consistent with Lewin's formulations, were tested:

1. Marginal Jews show social preferences more like Presbyterians than do nonmarginal Jews.

2. Marginal Jews show greater personal insecurity than either nonmarginal group.

3. Marginal Jews show less agreement with one another in their social preferences than either nonmarginal group.

4. Marginal Jews as individuals show less consistency in their social preferences than members of either nonmarginal group.

METHOD

Subjects

Data were collected at Stanford University in the spring of 1951. *Ss* were 40 marginal Jews, 40 nonmarginal Jews, and 40 Presbyterians of North European background. All were male students ranging in age from 17 to 37. All of the Presbyterian *Ss* indicated that both parents were Protestant and had English, Scotch, German, Dutch, or Scandinavian surnames.

In assembling the two Jewish groups, an effort was made to get in touch with all male students of Jewish background at the university, including agnostics. A questionnaire, entitled "Participation in Jewish Affairs," was used to select the marginal and nonmarginal Jewish samples, and was completed by 159 of these students. It contained 24 multiple-choice items covering synagogue and church membership and participation, religious education, religious festivals, Bible and prayer, dating and marriage, fraternities and resorts, food habits, charities, reading and radio listening, and general social contact. Each of the response alternatives had been weighted on a 7-point basis from −3 for extreme deviation from conventional Jewish norms to +3 for extreme adherence to such norms. Weights were assigned independently by 10 sophisticated Jewish men and women. For 82% of the 95 response alternatives the modal value contained a majority of the judgments. The average of the judged weights for any alternative, to the nearest integer, was assigned to that alternative. Thus to the question "Have you ever deliberately avoided social contacts with Jews?" a weight of −3 was assigned to the alternative Yes, and 0 to the alternative No. To the question "Thinking of the girls you have dated at college, what proportion would you say were Jewish?" a weight of +3 was assigned to the alternative "all have been Jewish," and −2 to the alternative "none have been Jewish." The range of possible scores for the whole scale was −27 to +34.

The split-half reliability of the scale was .85 by the Spearman-Brown formula. Some indication of its validity may be gained by noting that 18 out of 19 men selected by two rabbis (Orthodox and Conservative) as highly conventional in Jewish religious practice scored higher on a preliminary form than 92 out of 96 nominally Jewish members of the progressive American Veterans Committee in the same region.

For the present study the respondents scoring in the lowest quartile, with scores from −12 to −1, were retained as the marginal Jewish group. Those scoring in the upper quartile, with scores ranging from +11 to +29, were designated the nonmarginal Jewish group.

Procedure

By means of paired comparisons, *Ss* in each of the three groups were tested individually for social preferences, intra-group agreement, and intra-individual consistency with respect to 16 photographs. The photographs comprised four North European men, four Mediterraneans, four Jews, and four Negroes, with one professional, one white-collar worker, one skilled worker, and one unskilled worker in each ethnic set. Each of the full-face, black-and-white photographs was taken under standard flash lighting conditions, enlarged to standard size (head length about 5 cm.), cut away from its background, and mounted on a heavy buff 5 x 7 in. cardboard. These men were selected by going to places of work where members of the specified ethnic groups were likely to be found. Surnames and personally acknowledged identity were checked with the desired ethnic background in all cases except Negroes, where such a check was unnecessary. Their ages were between 20 and 35. Their attire, showing to the middle of the chest, was nearly uniform within each occupational level.

Photographs were presented as a series of 120 pairs. Each *S* was instructed to think of himself in a cafeteria near where he lives, and to choose a seat at a table for two, opposite either of the two men represented in the photographs that *E* had placed on the easel. It was indicated that no other seats were available in the cafeteria at the time, and that each of the two available places was equally far from *S*, who cannot tarry too long with his loaded tray. The home-town cafeteria situation was chosen because the face-to-face arrangement suggests more than an impersonal relationship, and the possibility exists of being observed by a friend or a family associate. Before each pair was presented, *S* was briefly reminded of this situation.

Each pair was presented once, in a fixed random order, with individual photographs counterbalanced for left-right position. *S* indicated which man he chose to sit with by pointing. Most *Ss* testified that the imagined context of the choices seemed real. They seemed to anticipate each choice with interest, were observed to fixate each picture at least once, made most of their choices within five seconds after presentation, and did not tire despite the long series of presentations.

At the conclusion of the paired comparisons, each *S* filled out a modified short form of the Knutson Personal Security Inventory, dealing with satisfaction with status, work, personal recognition, health, and general happiness. The seven-item short form shows a .85 correlation with the long form, whose validity is reported by Knutson (1952).

RESULTS

Social Preferences

Relative preferences for North Europeans, Mediterraneans, Jews, or Negroes within the ranges of stimulus difference represented in the photographs could be determined by summing the choices of photographs in each of these ethnic categories. The majority of such choices

would involve not only ethnic differences, but also occupational differences. In order to eliminate confounding ethnic with occupational considerations, only the 24 intra-occupational comparisons are considered, with ethnic preference scores obtained by summing choices of men in each of the four ethnic groups within these comparisons. In the same manner, occupational preference scores may be derived from the 24 intra-ethnic comparisons. Analysis of variance showed no significant differences among marginal Jews, nonmarginal Jews, and Presbyterians in the distribution of preferences among the four occupational levels.

Let us refer to the array of *S* groups involving marginal Jews, nonmarginal Jews, and Presbyterians as the "belonging" variable. Similarly, let us refer to the dimension involving North European, Mediterranean, Jewish, and Negro photographs as the "ethnic objects" variable. Analysis of variance shows that variance due to ethnic objects is sufficiently greater than the remainder variance (after removal from the total variance of variance due to ethnic objects and the ethnic objects × belonging interaction) so as to be significant at the .01 level.[1] The over-all test for hypothesized differences is the interaction between belonging and ethnic objects, which proved to be significant at the .01 level. In a separate analysis of variance involving only the preferences of marginal and nonmarginal Jewish *S*s on Jews and Mediterraneans, the belonging × ethnic objects interaction proved to be significant at about the .02 level.[2]

Fig. 1 shows means for each belonging class on each of the ethnic objects classes. Relative to the nonmarginal Jews, Presbyterians preferred North Europeans and Mediterraneans, and avoided Jews and Negroes. Marginal Jews tended to agree with nonmarginal Jews in avoiding North Europeans and favoring Negroes, but showed the same pattern of preferences as the Presbyterians with respect to Mediterraneans and Jews. That is, relative

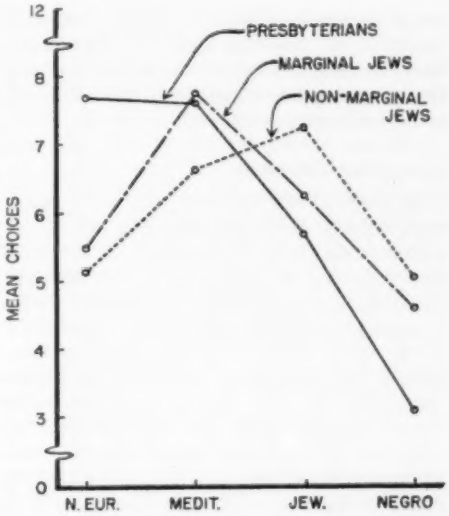

FIG. 1. MEAN CHOICES OF NORTH EUROPEAN, MEDITERRANEAN, JEWISH, AND NEGRO MEN BY MARGINAL JEWS, NONMARGINAL JEWS, AND PRESBYTERIANS IN 24 INTRAOCCUPATIONAL PAIRED COMPARISONS

to nonmarginal Jews, marginal Jewish *S*s tended to favor Mediterraneans and to avoid Jews. Each of these differences with respect to Mediterraneans and Jews is significant at the .05 level, the Presbyterian-nonmarginal Jewish difference with regard to Jewish ethnic objects being significant at the .01 level, by *t* test. Each of the differences with respect to North Europeans and Negroes is significant at the .01 level by *t* test.

Personal Security

No significant differences in personal security mean scores were found between groups of *S*s. Mean scores for nonmarginal Jews, marginal Jews, and Presbyterians were 7.35, 6.88, and 7.50 respectively in a possible range of 0 to 10. Corresponding *SD*s were 1.88, 1.54, and 1.66.

Among marginal Jewish *S*s a significant correlation ratio was found relating personal security and excess of choices of Mediterraneans over North Europeans (eta = 0.56, $P < .05$ for epsilon-square). Except for the extreme involving nearly total rejection of North European social objects relative to

[1] The paired comparisons technique fixes "between *S*s" variance including variance due to belonging at zero.

[2] Since only half of the intra-occupational comparisons are involved here, the "between *S*s" variance is not fixed at zero. This variance was removed from the total variance in figuring the "within *S*s" variance including the remainder variance.

Mediterraneans, higher personal security tended to be associated with preference for Mediterraneans. Neither nonmarginal Jews nor Presbyterians showed any such relationship in corresponding analyses.

Intra-group Agreement

Kendall's (1948) *u* gives a measure of agreement among judges making a given set of paired comparisons. This statistic is based on the proportion of judges for any comparison AB making the judgment "A preferred to B" as against the proportion saying "B preferred to A," and considers all comparisons. Considering the average of intra-occupational *u*'s for each of the three groups of *S*s, over-all *u* for nonmarginal Jews was .10, for marginal Jews was .11, and for Presbyterians was .22. Using a related statistic given by Kendall which is distributed as chi square, each of these three *u*'s proves to be significantly different from zero at the .001 level. Converting the three chi squares to *Z*s, the difference between coefficients of agreement for marginal Jews and nonmarginal Jews proved to be not significant, while the difference between each of these groups and the Presbyterians was significant at the .001 level.

Intra-individual Consistency

Among the six paired comparisons at any occupational level, social preferences indicated by any *S* may be completely consistent (transitive) or may involve some inconsistency or intransitivity. For example, if we label the four professional men A, B, C, and D, the judgments "A preferred to B," "B preferred to C," and "C preferred to A" are intransitive and constitute a circular triad. The number of circular triads of photograph pairs was determined for each *S*, adding together the number of circular triads from each occupational level. No circular triads at all represents perfect consistency for an *S*.

Twenty-eight Presbyterians, 23 marginal Jews, and 18 nonmarginal Jews showed perfect consistency, the remainder showing between one and five circular triads out of a possible eight. The frequency of perfect consistency among Presbyterians was significantly greater than that for the nonmarginal Jews, at the .02 level by chi square. The intermediate frequency of the marginal Jews was not significantly different from either of the other groups.

DISCUSSION

The first hypothesis given in the introduction stated that marginal Jews would show social preferences more like Presbyterians than would nonmarginal Jews. This hypothesis is clearly supported by the data. But the patterns of preferences of marginal Jews and members of the dominant group are by no means identical. While nonmarginal Jews tend to orient themselves toward Jews, and Presbyterians toward both North Europeans and Mediterraneans, marginal Jews tend to orient themselves toward Mediterraneans—but *not* toward North Europeans. In other words, the data suggest that marginal Jews may identify themselves (or prefer to associate or be identified) with a group which is not the dominant and perhaps unaccepting ethnic group, but which is just as acceptable to members of the dominant group as the dominant group itself. Since this "compromise" group is on the whole physiognomically more similar to the Jewish minority than the dominant group, visual "passability" may also be a determining factor in orientation.

The second hypothesis predicted that marginal Jews would show greater personal insecurity than either of the nonmarginal groups. The data fail to support this hypothesis. The recent research of Kerckhoff and McCormick (1955) and Mann (1958) suggests that this finding is to be expected if marginal Jews do not orient themselves toward the dominant North European group. In the Kerckhoff and McCormick study of 54 Chippewa Indian children, high insecurity scores were found only among those *S*s who were sufficiently Indian-looking to be rejected by whites *and* who identified with white culture. If only one or if neither of these conditions applied, personal insecurity was lower, i.e., about the same as of white schoolmates.

With adult members of the South African Colored (mulatto) population, Mann's results were similar. He studied 25 high passable (near white) men and women and a matched group of 25 who were low passable. In addition to measures of orientation toward whites and of "marginal personality" traits (insecurity, self-pity, and sensitivity), Mann obtained a

measure of "barrier acceptance"—acceptance of restrictions applied by whites to Colored-white interaction. He found no correlation between passability and the personality scores. But he found two sets of conditions under which personal insecurity (and associated personality measures combined into one score) was high: (*a*) low passability, low barrier acceptance, *and* a high "undecided" score on the white-Colored orientation measure; and (*b*) high passability, low barrier acceptance, *and* pro-white orientation.

Knowing that the marginal Jews in the present study tended not to orient themselves toward the dominant North Europeans, it may be expected in the light of the two studies just reviewed that they would not on the average show greater insecurity than the other groups of *S*s. The relationship between security and North European vs. Mediterranean orientation, showing greatest personal security for marginal Jews who moderately favor Mediterraneans and progressively less security as a preference for North Europeans increases, provides supporting evidence for the orientation-insecurity hypothesis, with orientation as a continuous variable. It is probable that Sarnoff's (1951) anti-Semitic Jews were so strongly oriented toward North Europeans as to "identify" with them and reject themselves as Jews by anti-Semitic criteria. Some marginal Jews in the present study might possibly be so characterized, but evidently most of them cannot: it is apparent that negative orientation toward the Jewish minority does not necessarily imply positive orientation toward the dominant North European group. It is reasonable to suppose that marginal Jews see the North European group as threatening or at least unaccepting, while the Mediterranean group is seen as more accepting and otherwise more permeable.

Results with respect to the third and fourth hypotheses are not surprising in view of the findings already discussed. Contrary to the deduction from Lewin's theory, marginal Jews are no less agreed about in-group and out-group preferences than nonmarginal Jews, even though Presbyterians are more in agreement. It is reasonable to suppose that identification with some non-Jewish group—whether ethnically defined, as suggested by the preference for Mediterraneans, or nonethnically defined—could maintain agreement among marginal Jews at least as great as among nonmarginal Jews. It might alternatively be argued that Jews, whether marginal or nonmarginal, are generally less prejudiced toward out-groups than members of the dominant ethnic group, or at least lack uniform ethnic standards of social judgment. The only data that might be advanced from this study in support of this argument is that the social preference curves of the two Jewish groups (Fig. 1) were flatter than that of the Presbyterian group.

The fourth hypothesis dealt with intra-individual inconsistency in preferences to be expected from assumed tension states. The data fail to show that marginal Jews are any more inconsistent than either of the nonmarginal groups. These results agree with the finding that marginal Jews do not differ from either of the other groups of *S*s in personal insecurity (tension). The significantly greater inconsistency shown by individuals in the nonmarginal Jewish group compared to the Presbyterian group, however, is curious. It is possible that nonmarginal Jews were less interested than Presbyterians in the social objects they were choosing among, but there is no evidence for this. It is possible that nonmarginal Jews are less rigid in their standards of interpersonal preference than Presbyterians, shifting standards from one pair of faces to another. It is also possible that the social environment of nonmarginal Jews may be somewhat less differentiated to begin with, relative to the other two groups under consideration, and that recall of recently expressed preferences is therefore less exact than among the other groups. No relevant evidence can be offered, however, at this time.

In general, it may be concluded that marginal Jews do not necessarily orient themselves toward the dominant ethnic group of the society, and that tension derived from unwilling nominal membership in an underprivileged group may be reduced by identifying or being associated with a nonthreatening alternative group. It may prove valuable in the future to investigate availability of new groups or their norms to marginal individuals as a factor in marginal personal dynamics distinct from barriers of prejudice which inhibit social mobility.

SUMMARY

Forty marginal Jews, 40 nonmarginal Jews, and 40 Presbyterians of North European background individually made paired comparisons of unlabeled photographs of 16 men of four ethnic groups. Selection of one photograph of any pair meant preference for that man as a luncheon partner. Presbyterians as a group preferred North Europeans and Mediterraneans, nonmarginal Jews preferred Jews, and marginal Jews preferred Mediterraneans. Nonmarginal Jews showed less agreement in their choices than Presbyterians, and marginal Jews showed an intermediate level of agreement. Both marginal and nonmarginal Jews showed less intra-individual consistency in their choices than Presbyterians, but marginals did not show less consistency than nonmarginal Jews. A measure of personal security failed to show differences between the three groups. However, marginal Jews who preferred North Europeans to Mediterraneans tended to be less secure than those who moderately favored Mediterraneans.

The results suggest that marginal Jews tend to identify themselves with non-Jewish groups which do not reject them, and consequently do not maintain the tension which is associated with identifying with an aggressor.

REFERENCES

GLASER, D. Dynamics of ethnic identification. *Amer. J. Sociol.*, 1958, **23**, 31–40.

KENDALL, M. G. *Rank correlation methods.* London: Charles Griffin, 1948.

KERCKHOFF, A. C., & MCCORMICK, T. C. Marginal status and marginal personality. *Soc. Forces*, 1955, **34**, 48–55.

KNUTSON, A. L. Personal security as related to station in life. *Psychol. Monogr.*, 1952, **66**, No. 4 (Whole No. 336).

LEWIN, K. *Resolving social conflicts.* New York: Harper, 1945.

MANN, J. W. Group relations and the marginal personality. *Hum. Relat.*, 1958, **11**, 77–92.

RADKE-YARROW, MARIAN, & LANDE, B. Personality correlates of differential reactions to minority group-belonging. *J. soc. Psychol.*, 1953, **38**, 253–272.

SARNOFF, I. Identification with the aggressor: Some personality correlates of anti-Semitism among Jews. *J. Pers.*, 1951, **20**, 199–218.

STECKLER, G. A. Authoritarian ideology in Negro college students. *J. abnorm. soc. Psychol.*, 1957, **54**, 396–399.

(Received August 6, 1958)

REACTIONS TO UNFAVORABLE PERSONAL EVALUATIONS AS A FUNCTION OF THE EVALUATOR'S PERCEIVED ADJUSTMENT[1]

EDWARD E. JONES, STEPHEN L. HESTER, AMERIGO FARINA, AND KEITH E. DAVIS
Duke University

THE perceiver who is faced with a complex pattern of responses from another person, and is asked to integrate and express his first impression, is faced with a difficult and challenging task. Clearly, the resulting impression depends to an important extent on the content of the observed behavior, but it also depends on the interaction context in which the particular behavior occurs. Jones and Thibaut (1958) have attempted to classify various common types of social interaction and to make tentative predictions concerning the cognitive effects of interactions having different interpersonal goals. Their basic position is that various *inferential sets* are aroused or made salient by cues defining the type of interaction taking place, and that these sets partially determine the selection and integration of information about the participants in the interaction.

An obvious or implicit goal in many interactions involves the maintenance of one's values and self-image. While empirical data are not abundant in support of the point, it seems obvious that a perceiver's interpretation of an act and his inferences about the source of the act vary with the degree to which the stimulus person's behavior directly affects the perceiver and thwarts or promotes his values and personal integrity (cf. Jones & de Charms, 1957). The present study is explicitly concerned with such variation, since it deals with a situation in which one member of a pair of *S*s is criticized and belittled by a stimulus person (SP) while the other member observes. Subsequent ratings of the SP by both the "involved" *S*s and the "bystander" *S*s serve as the major dependent variable in the experiment.

[1] This study was formulated in the program of the Organization Research Group of the Institute for Research in Social Science at the University of North Carolina. The O.R.G. is supported by the Office of Naval Research [Nonr-855(04)]. The research was greatly facilitated by grants to the senior author from the Ford Foundation and the Duke University Research Council.

Given such a situation, consider the plight of the involved *S*, i.e., the person who serves as the target of criticism and derogation. He may respond ideally in one of the following ways: (*a*) he may view his rating task as an opportunity to retaliate in kind, thus assigning the derogator more negative ratings than does the bystander; (*b*) he may accept the derogation as justified, and rate the derogator as positively as or more positively than the bystander (such a strategy might involve some degree of distortion of the derogation attempt); or (*c*) he may reject the derogation but relieve the derogator of responsibility for his act, i.e., in some way dissociate the source from his act. In the latter case, the reaction of the involved *S* would either be neutral or positively sympathetic, depending on the means by which dissociation is achieved.

In the present experiment the derogation included comments especially intended to reduce the frequency with which the second (acceptance) reaction was likely to occur. Thus, the derogation was quite arbitrary and difficult to justify. Separate treatments were also designed to promote the first (hostile) or the third (dissociative) reaction in different involved *S*s. These treatments differed only in the kind of background information provided about the SPs. Half of the *S*s were led to believe by prior information that the derogator was a stable, emotionally secure person; the other half learned that the derogator was maladjusted and lacking in insight.

This particular variation was chosen because it was felt that the involved and bystander *S*s would use the information about adjustment in different ways. The major experimental hypothesis was that the involved *S* would respond with less hostile ratings than the bystander when the derogator was maladjusted, and with more hostility than the bystander when the derogator was well adjusted.

To appreciate the rationale of this prediction, it is important once again to examine the

position of the involved *S*. When he is attacked by the well-adjusted derogator, it is difficult to take refuge in the thought that the derogator is highly inaccurate, a victim of projection, and so forth. Faced with the maladjusted derogator, however, he may console himself with the thoughts that (*a*) the derogator's judgments of others are distorted since they are colored by his own problems, and (*b*) the derogator is not fully responsible for the act of judgment, being a victim of a traumatic past history. Therefore, because the SP's perceived maladjustment affords the involved *S* the opportunity of reducing the sting of the derogation, his tendency to blame the SP should be held in check by feelings of sympathy.

By comparison, the bystander is expected to follow the tendency observed in other experiments (e.g., Jones & deCharms, 1958) to like the well-adjusted and dislike the maladjusted SP regardless of his derogatory behavior. It is assumed that the combined information, maladjustment and derogation, summate for the bystander to produce a very negative impression, whereas the perception of maladjustment by the involved *S* carries positive overtones because of its functional significance and tends to cancel out the hostility that would otherwise be aroused by the derogation.

This main hypothesis, and others associated with the same basic rationale, will be elaborated in presenting the results.

METHOD

Subjects

The *S*s were all female and consisted of 50 freshman and sophomore nursing students from Duke University. All *S*s were volunteers, though they were paid for their services in the experiment. *S*s were always run in pairs, one *S* being involved in the derogation, the other being a bystander. No effort was made to control for the degree of acquaintance between members of each experimental pair, but all *S*s stated that they were well acquainted and many were roommates. The fact that *S*s were well acquainted would, if anything, tend to reduce the power of the "involvement" variation, since the bystander *S*s were also sympathetically or vicariously involved to some degree.

Procedure

The chronology of experimental events may be outlined as follows:

1. Each pair of *S*s was introduced by *E* to a confederate posing as the instructor of a "senior course in personality assessment." The "instructor" then described two features of his course: (*a*) each student was himself clinically assessed early in the course and later (*b*) was given an opportunity for brief observation of an unknown *S* to learn "the possibilities and the limitations of first impression formation." One of the *S*s was to help in this latter requirement by serving as a "guinea pig" for two students observing through a one-way mirror. The other was to serve as an "innocent bystander." Allegedly to provide some feedback for the former *S* as to how she struck others, and to make the experiment more interesting for both *S*s, they were invited to stay and overhear the students discuss their first impressions.

2. The *S*s were then given a summary data sheet containing background information on each of the two student-observers soon to arrive behind the mirror. These sheets were presented as summaries of the above mentioned clinical assessment. The *S*s were to read these sheets while the "instructor" was presumably rounding up the two students. The background data sheet on one of the student observers emphasized her unstable home life, inadequate emotional resources, and underlying anxiety. In short, the data sheet was constructed to convey the impression of a relatively maladjusted college girl. The other background data sheet portrayed a student who was effective, well-rounded, insightful, and whose home life had been happy and rich with support and affection. For one group of *S*s (the *derogator-mal* group), the data sheets presented the SP who did not derogate as well-adjusted. For a second group of *S*s (the *derogator-well* group), the background data sheets were simply reversed.

3. After the *S*s read the background sheets, *E* openly flipped a coin to decide which of the *S*s would be evaluated in the experiment (the involved *S*) and which would be the bystander. The bystander was instructed to sit and observe carefully, and to attempt to decide what impression the involved *S* would make on her if she had never seen her before. The involved *S* then, in sequence, (*a*) sat and faced the one-way mirror for about 30 sec., (*b*) wrote a short sentence on the blackboard, and (*c*) read a few standard passages of prose and poetry while sitting and facing the mirror.

4. Both *S*s then filled out a rating scale under instructions to predict the ratings which they were soon to hear, via the loudspeaker, from the student observers (SPs).

5. After predicting how the involved *S* would strike the SPs, both *S*s listened to a standard tape recording which they were led to believe was an interview then taking place between the "instructor" and two of his (female) students who had just observed and rated the involved *S*. The trained actresses representing the two SPs reported their ratings and made relevant comments in a rather informal conversation with the "instructor." One rated the involved *S* in a derogatory manner on certain key items and rather arbitrarily expressed her irritation with the *S*. Her final comment was patronizing and provoking: "I think she's about average for a nursing student." The other SP was consistently more hesitant in her judgments and showed greater warmth and sympathy. In order to simulate a concern with anonymity, the SPs referred to each other as Miss Red

and Miss Blue and were so addressed by the "instructor."

6. After the tape recording had finished, the *S*s were asked to rate the two SPs on a number of items. Finally, the *S*s were asked to fill out a postexperimental questionnaire and the actual details of the experiment were fully disclosed to each *S*. There was no evidence that any of the *S*s were aware of the true nature of the experiment, and all expressed surprise that the interview was tape-recorded.

Experimental Design

The two independent variables of the experiment were (*a*) whether the derogator was presented as well-adjusted or maladjusted, and (*b*) whether the *S* was being evaluated or only observing. The combination of the two produced four cells of *S*s in a two-by-two design: 12 involved and 12 bystander *S*s in the derogator-mal condition, and 12 involved and 14 bystander *S*s in the derogator-well condition. Cell numbers were unequal because three involved *S*s and one bystander *S* in the derogator-well group quite obviously confused the two stimulus persons.

Rating Scales and Questionnaires

Two separate rating scales, of quite different form, were used at different points in the experiment. The first scale was employed by each *S* in making predictions of the impression created by the involved *S*. The same scale was presumably to be used by the *SP*s as a basis for their discussion. This scale consisted of 16 items with eight response categories between two polar adjectives like "extremely self-centered" and "extremely altruistic." Other items covered such qualities as popularity, tolerance, generosity, self-confidence, anxiety, etc. Since all items were highly evaluative in nature, a summary prediction score could be obtained by combining the different item values into one total score. A high score would mean that the rater expected a favorable reaction to the involved *S*. For the involved *S*, of course, this prediction was a form of self-rating.

The main dependent variable was another rating scale specifically constructed to reflect the *S*s' impressions of the two *SP*s. This scale consisted of 13 trait pairs (i.e., 26 items) which measured such perceived characteristics as likability, adjustment, dependability, and hostility. The ratings were made on a six-point scale reflecting the degree of agreement with simple declarative statements about the stimulus person being rated. The scale was constructed in such a way that a positive statement like: "As a person, she is extremely likable" was in all cases matched by its negative counterpart in a different item such as "I find it hard to like this person to any extent." The results are always discussed in terms of item pairs, instead of items, in order to control for response set.

By reversing the negative item score and comparing this with the positive score for that item, it is possible to obtain a measure of reliability for each pair. With the exception of one pair of items, the reliability was uniformly high. Forty-three per cent of the positive item scores were identical with the reversed negative item scores (with a range for each item of from 32 to 54%), and 86% of the pair scores were divergent by no more than one point (with an item range of from 72 to 93%). The exception was a pair of items dealing with suggestibility which, in retrospect, seems ambiguously worded. Results on this item pair are not included in the present report.

The postexperimental questionnaire consisted of six questions about the factors that the *S*s had consciously considered in making their ratings of the two SPs. Three of the items were about Miss Red and three were about Miss Blue. The most important question listed four factors which might have influenced the *S*s and asked the *S*s to rank these items in terms of relative influence.

RESULTS

Validation of Attempted Inductions

In experiments of this type the stimulus conditions can be constructed by, but not entirely defined by, the investigators themselves. It is important to assess whether the *S*s perceived the stimulus variations in line with experimental intentions.

Perceived differences between maladjusted and well-adjusted stimulus persons. Separate analyses of variance were run on each trait pair to assess the effects of *S* involvement and SP adjustment on ratings of the derogator. These analyses were repeated on ratings of the nonderogator. In each case the treatment cells were equated at 12 by omitting two *S*s in the bystander, derogator-well condition. The results of these analyses show that the well-adjusted derogator, relative to the maladjusted derogator, was attributed better judgment, more insight, and was seen as more well-adjusted, dependable, and a better judge of persons. Very similar results obtain for the nonderogator. These results clearly establish the fact that the *S*s read and assimilated the background information presented to them.

Perceived differences between derogator and nonderogator. If the tape recording scripts were effectively written and the roles successfully enacted, then the derogator should be judged relatively hostile and conceited by all *S*s. The analyses reveal this to be the case, whether the derogator was well-adjusted or maladjusted. Thus the derogator was seen as more "superior and pretentious," more "hostile and antagonistic," and basically "cold and aloof." The impact of the interview comments, then, seems to coincide closely with the experimental intention.

TABLE 1

JOINT EFFECTS OF SP ADJUSTMENT AND *S* INVOLVEMENT ON "LIKABILITY"

(Based on discrepancy between ratings of derogator and nonderogator)[a]

	Condition	
	Derogator-mal	Derogator-well
Involved *Ss*		
$\bar{X}$	3.92	2.17
SD[b]	1.78	2.29
Bystander *Ss*		
$\bar{X}$	4.33	−.57
SD	3.17	3.96

[a] The larger the mean value, the greater the preference for the nonderogator.

[b] The four variances are not significantly heterogeneous.

TABLE 2

CORRELATIONS BETWEEN "LIKABILITY" AND "ADJUSTMENT"

(By adjustment condition, derogator only)

	Condition	
	Derogator-mal	Derogator-well
Involved *Ss*	.23	.48
Bystander *Ss*	.53	.81

Major Experimental Hypotheses

Hypothesis 1. The involved Ss will show greater liking for the maladjusted derogator than the bystander Ss, and less liking for the well-adjusted derogator than the bystander Ss. The rationale for this hypothesis has already been outlined in terms of the involved *S*'s greater need to dissociate the act of derogation from its source. When dissociation may be easily achieved, in the derogator-mal condition, a sympathy reaction is predicted. When dissociation is not easily achieved, in the derogator-well condition, a reaction of hostile retaliation is predicted. The bystander should be less concerned with reducing the sting of derogation and should be more inclined to like the well-adjusted more than the maladjusted SP in general.

In order to test this prediction, likability ratings of the nonderogator were subtracted from likability ratings of the derogator to provide a discrepancy score for each *S*. In this way, ratings of the nonderogator served as a control for generalized response biases in using the rating scale. The resulting scores (see Table 1) were examined by a t test for comparison of two mean differences (Walker & Lev, 1953, p. 159). Thus the mean score for the involved *Ss* in the derogator-mal condition was subtracted from the mean score of other involved *Ss* in the derogator-well condition and the difference compared with the same difference obtained from the bystander *S* means. The resulting value of t confirms the prediction, but not strikingly so ($t = 1.87$; $P < .05$, one-tailed test). The difference between involved and bystander *Ss* is significant for the derogator-well condition ($t = 2.11$; $P < .025$), but not for the derogator-mal condition ($t = .39$) (one-tailed tests in each case). Nevertheless, the means are all in a direction consistent with the first hypothesis.

Hypothesis 2. The bystander Ss will tend to associate likability with good adjustment to a greater extent than the involved Ss in rating the derogator. Essentially a corollary of the first hypothesis, this prediction is concerned with comparing correlations between likability ratings and adjustment ratings of the derogator in each adjustment condition. It is assumed that the general tendency for these two characteristics to be seen as related will be inhibited by the involved *Ss*' desire to attribute the derogation to an unreliable or distorting source. If this end can be achieved by perceiving the derogator as maladjusted, the involved *S* should not express as intense dislike toward the derogator.

The relevant correlations are presented in Table 2. The differences are in the right direction in each condition, but do not reach statistical significance in either.

Hypothesis 3. When the derogator is maladjusted, both involved and bystander Ss will be more "sensitive" to background information than to the remarks of the derogator. When the derogator is well-adjusted, the bystander Ss will again be more sensitive to background information, but the involved Ss will be more sensitive to the remarks. This prediction again stems from the basic rationale supporting Hypotheses 1 and 2. A first specific assumption is that the bystander is less concerned with remarks concerning another person than she is with the personality and social history of the SPs. A second assumption is that the involved *S* needs to emphasize the maladjustment of the derogator in order

to achieve the dissociation discussed above. When the derogator is well-adjusted, however, concern with the SP's background merely points up the painful fact that the derogation cannot be discounted as a projective distortion. In this condition, therefore, the involved *S*s should show a general tendency to avoid the background information relative to the bystander *S*s and show greater sensitivity to the remarks of the derogator when forced to choose between remarks and background.

The sensitivity of the *S*s to different sources of data about the SPs was inferred from the postexperimental questionnaires in which each *S* was asked: "Which of the following kinds of information was most important in determining how you felt about Miss Red? (Rank in order of importance)—(*a*) her background and family life; (*b*) her Rorschach scores and the instructor's comments; (*c*) the quality of her voice; and (*d*) the things she said about the nursing student who performed." The hypothesis was tested by classifying each *S* in terms of whether Item (*a*) or (*d*) was considered more important, and testing the association between this classification and that into involved and bystander by chi square. As Table 3 indicates, the hypothesis is confirmed. In the derogator-well condition, the involved *S*s paid greater attention to the remarks (or less attention to the background) than did the bystander *S*s ($P < .05$); in the derogator-mal condition all *S*s paid more attention to the background information. Stated differentially, the involved *S*s showed greater interest in the remarks of the derogator when she was presented as well-adjusted than as maladjusted. The bystander *S*s show no such tendency, always expressing greater interest in the background information.

TABLE 3

ASSOCIATION BETWEEN INVOLVEMENT AND SENSITIVITY TO REMARKS VERSUS BACKGROUND INFORMATION

(By adjustment condition, derogator only)

	Condition			
	Derogator-Mal		Derogator-Well	
	R[a]	B[b]	R[a]	B[b]
Involved *S*s	4	8	10	2
Bystander *S*s	3	9	5	9
	$\chi^2 \sim .00$		$\chi^2 = 4.21$[c]	

[a] Ranks "the things she said about the nursing student who performed" as more important than "her background and family life."

[b] Ranks "her background and family life" as more important than "the things she said . . ."

[c] Corrected for continuity. $P < .05$, two-tailed test.

Hypothesis 4. The involved Ss will see the derogator as more maladjusted than the nonderogator; the bystander Ss will show this tendency to a smaller extent. This again is in keeping with the theme of dissociation which presumably characterizes the involved *S*s. The hypothesis is not confirmed. When the SP is maladjusted, the differences are in the right direction but do not reach significance. That is, the involved *S*s saw the (maladjusted) derogator as more maladjusted than the (maladjusted) nonderogator to a significant extent ($P < .05$, by t test) and the difference for the bystander *S*s does not come close to significance. However, the difference between these differences is not significant.

When the SP is presented as well-adjusted, the differences are actually in the opposite direction. Both involved and bystander *S*s saw the derogator as more maladjusted than the nonderogator, but this difference was slightly greater for the bystander *S*s.

Relationships Involving Prediction Ratings

It is a commonplace that complex patterns of social stimulation do not elicit uniform reactions in all who are exposed to them. There are inevitable individual differences in the assimilation and interpretation of social events. While the basic strategy in the present investigation was to establish sufficient control over the *S*s' behavior to yield meaningful generalizations about the "average person," a provision was made to examine one dimension of individual differences. Each involved and bystander *S*, it will be recalled, made specific predictions as to how the involved *S* would strike the two SPs. While both sets of predictions are of interest in providing information about the *S*s' expectations, those of the involved *S*s are particularly interesting since they are presumably related to self-perception and affected by the need to maintain a positive self-image.

Because of the ambiguous status of the prediction ratings, specific hypotheses were not formulated in advance. It was generally expected that the *S*s who predicted unfavorable

TABLE 4
CORRELATION BETWEEN PREDICTION RATINGS[a] AND "LIKABILITY"
(By adjustment conditions, derogator only)

	Condition	
	Derogator-mal	Derogator-well
Involved *Ss*	−.616	+.835
Bystander *Ss*	−.057	+.246

[a] As explained in the text, the prediction ratings were combined by summation to provide an index of the degree to which a positive evaluation was expected from the SPs. A positive correlation means that a prediction of favorable evaluation goes with liking the derogator better.

ratings would consider the derogator a better judge of personality than those who predicted favorable ratings. This was clearly borne out for both involved and bystander *Ss* ($P < .01$ by analysis of variance). Beyond this, however, our only hope was to make *ad hoc* sense out of the findings involving the prediction ratings.

The most striking relationship is that between prediction rating and "likability" in the two stimulus conditions. The product-moment correlations presented in Table 4 reveal a dramatic contrast in the direction of correlation for the involved *Ss*, when the derogator-mal is compared with the derogator-well condition. When the derogator was maladjusted, the *Ss* who predicted an unfavorable evaluation of self-expressed greater liking for the derogator than did those who predicted a favorable response. When the derogator was well-adjusted, however, those who predicted an unfavorable evaluation expressed *less* liking for derogator than the favorable self-raters.

The reliability of this finding with the involved *Ss* is too high to be ignored, but interpretation rests on the meaning of rating oneself low versus rating oneself high on the prediction ratings. One thing is clear: those *Ss* who rated themselves low more closely conform to the results anticipated for the involved *Ss* in general than do the high self-raters. In fact, the high self-raters behaved in most respects like the bystander *Ss*. From this it may be inferred that the low self-raters were more "involved" in the experiment than were the high self-raters, and that low predictions served as a protective device for the person who was uncertain about her impact on others. Through this mechanism of self-derogation the low raters may have prepared for the worst. It would then follow that the lows should be more sensitive to the cues of maladjustment in the derogator than are the high self-raters, and this tends to be the case. In the derogator-mal condition, five out of six lows (with prediction ratings below the median) showed more interest in the derogator's background than in the remarks she made. Only three out of six highs showed more interest in the background. Both highs and lows were more sensitive to the remarks when the derogator was well-adjusted.

Regardless of the validity of the interpretation offered, it is clear that one's perception of the impression he creates in others is an important source of variance when attempting to predict his response to criticism and derogation. Furthermore, it is not simply a matter of responding to the correctness or incorrectness of one's prediction. Otherwise, there would be no differences between involved and bystander *Ss*. The fact that the derogator's adjustment is such a critical factor further suggests the plausibility of interpretations referring to anxiety about the self and related tactics of ego-defense.

DISCUSSION

In the present experiment, different contexts of interaction between perceivers and stimulus persons were established by varying the involvement of the perceivers, or the relevance to the perceivers of the stimulus persons' behavior. By also varying the alleged prior adjustment of the stimulus persons, it was possible to show the effects of the different interaction contexts on the selection and utilization of different perceptual cues. Thus the involved and bystander *Ss* differed in their inferences from and integration of the objectively identical cues with which they were provided in a given experimental treatment.

The perceptual differences that exist between *Ss* as a function of their roles in the situation are consistent with Jones and Thibaut's (1958) theoretical formulation, which emphasizes the cognitive consequences of different interaction goals and the *inferential sets* that are aroused by such goals. The present experiment is not directly concerned with comparing the consequences of arousing different inferential

sets, as was that by Jones and deCharms (1958). It was intended, rather, to throw light on how *S*s who are involved, i.e., who perceive the goal of the interaction as that of maintaining their self-concept and integrity, utilize social information to achieve this goal. Such *S*s may be said to be operating under a *value maintenance set*, whereas the bystanders were not (at least to the same degree). The differences between involved and bystander *S*s in their perception of the stimulus persons may be reasonably attributed to the differential degree of threat to the self-concept. Each group of *S*s draws different inferences from the information concerning maladjustment, and the impressions of the involved *S*s show an attempt to utilize this information in such a way as to maintain or regain their personal prestige. Those involved *S*s who expected unfavorable evaluations from others seemed particularly likely to behave in this way. It is inferred that these *S*s were especially anxious and uncertain about the impression they were making on others, and that they were therefore quite defensive.

All *S*s tended to see the well-adjusted stimulus person as more well-adjusted and the maladjusted person as more maladjusted. All *S*s tended to attribute more hostility to the derogator than to the nonderogator. But integration of these two types of information differed as a function of involvement. As predicted, the involved *S*s tended to like the derogator less than the bystander *S*s when the derogator was well-adjusted, but slightly more than the bystander *S*s when the derogator was maladjusted. To be sure, the "second-order" prediction was confirmed largely because of the differences when the derogator was well-adjusted, but the over-all consistency of the results and the fact that the bystander *S*s were not altogether *un*involved in the derogation increase one's confidence in the reliability of this main finding.

The present results may, of course, be viewed in a number of different theoretical contexts. One appropriate way of looking at the findings is to see the involved *S* as attempting to reinstate the equilibrium that was upset by the derogatory remarks. Heider (1946) points out that one way of doing this is to attribute the negative act to a negative person, i.e., to devalue the person who is the source of the negative act. This is essentially the predicted and obtained result in the derogator-well condition. Horowitz, Lyons, and Perlmutter (1951) experimentally verified Heider's statement, but pointed out that another way of reinstating equilibrium would be to dissociate somehow the act from the source. It would then be possible to dislike the act and still like the source. Such dissociation was predicted in the derogator-mal group (i.e., the involved *S*s liking the derogator more than the bystander *S*s), but did not reach statistical significance. Apparently, the general tendency to dislike maladjusted people was powerful enough to inhibit the tendency to dissociate act from source.

There is also a clear relationship between the present experiment and a study by Harvey, Kelley, and Shapiro (1957), who propose three possible reactions to unfavorable evaluations of the self: distortion, dissociation, and devaluation. In the present study no measure of distortion was obtained, but it is possible that distortion (i.e., misperceiving the amount of derogation) is related to the *S*s' expectations of ratings received. Thus it may be that those involved *S*s who predicted that they would be positively evaluated distorted the actual derogation more than those who predicted an unfavorable reaction. Such a relationship would help explain the observed differences between high prediction and low prediction *S*s. That is, the high prediction *S*s seem to have been less disturbed by the derogation in both maladjusted and well-adjusted conditions.

SUMMARY

Pairs of *S*s listened to two stimulus persons, allegedly conversing in an adjacent room, evaluate the personality of one member of the pair. One of the stimulus persons made derogatory remarks about the "involved" member, the other was more noncommittal and mildly sympathetic. In one experimental condition the derogator was pre-identified as being maladjusted, the nonderogator as well-adjusted. In another condition the identifications were reversed. After the *S*s had heard the evaluations, they rated the stimulus persons on a number of traits, and listed the factors that

had been most important to them in forming their impressions of the stimulus persons.

The main results were as follows:

1. In line with the basic prediction, the *S* who served as the target of derogation liked the maladjusted derogator better than the innocent bystander did and like the well-adjusted derogator less. Only the latter finding was significant by itself.

2. Also as predicted, relative to expressed interest in the background information, the involved *S* paid more attention to the remarks concerning her when the derogator was well-adjusted than when she was maladjusted. No such difference was observed in the other *S*s.

3. Those involved *S*s who expected to be unfavorably evaluated showed in rather extreme form the behavior predicted for all involved *S*s, liking the maladjusted derogator more and the well-adjusted derogator less than the *S*s who expected a favorable evaluation from the stimulus persons.

The results were generally consistent with the rationale and hypotheses of the study. Among other things, they point up the importance of considering the particular social context in which the behavior of a stimulus person occurs. They also are in keeping with a theory of cognitive balance or equilibrium.

REFERENCES

HARVEY, O. J., KELLEY, H. H., & SHAPIRO, M. M. Reactions to unfavorable evaluations of the self made by other persons. *J. Pers.*, 1957, **25**, 393–412.

HEIDER, F. Attitude and cognitive organization. *J. Psychol.*, 1946, **21**, 107–112.

HOROWITZ, M. W., LYONS, J., & PERLMUTTER, H. V. Induction of forces in discussion groups. *Hum. Relat.*, 1951, **4**, 57–76.

JONES, E. E., & DECHARMS, R. Changes in social perception as a function of the personal relevance of behavior. *Sociometry*, 1957, **20**, 75–85.

JONES, E. E., & DECHARMS, R. The organizing function of interaction roles in person perception. *J. abnorm. soc. Psychol.*, 1958, **57**, 155–164.

JONES, E. E., & THIBAUT, J. W. Interaction goals as bases of inference in interpersonal perception. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior*. Stanford: Stanford Univer. Press, 1958, Pp. 151–179.

WALKER, HELEN M., & LEV, J. *Statistical inference*. New York: Holt, 1953.

(Received August 6, 1958)

TEST OF A MODEL FOR OPINION CHANGE[1]

NORMAN H. ANDERSON[2]
Yale University

EXPERIMENTS in persuasion generally find that the greater is the advocated change in opinion, the greater is the change produced. Such results have been obtained, for instance, by Goldberg (1954), Fisher, Rubinstein & Freeman (1956), Hovland & Pritzker (1957), and Fisher & Lubin (1958). These investigators have also considered the possibility that the observed change is a linear function of (proportional to) the change advocated, and their results give some, not unequivocal, support for this assumption.

These considerations have been formalized and applied to the primacy-recency problem in a recent article (Anderson & Hovland, 1957). Specifically, it was assumed that the opinion before and after the presentation of a communication were linearly related, according to the equation

$$X_1 = X_0 + S(C - X_0) \qquad [1]$$

where

X_0 is the opinion before presentation of the communication
X_1 is the opinion after presentation of the communication
C is the fixed point of the communication
S is a coefficient of proportionality

The quantities, C and S are called the parameters. The fixed point, C, may be considered as the position of the communication on the opinion continuum. Operationally, C is defined as that initial opinion, X_0, which is left unchanged by the presentation of the communication (regardless of the opinion ostensibly advocated by the communication). The coefficient, S, may be thought of as the susceptibility of the person to the communication. It is analogous to the Efficiency Index of Hovland, Lumsdaine, and Sheffield (1949), and to the Conformity score of Fisher, Rubinstein, and Freeman (1956).

From Equation 1, the change in opinion may be written as $S(C - X_0)$, from which it can be seen that the greater is the value of S, the greater is the change in opinion in the direction of the communication. For simplicity, it will be assumed here that the values of S lie between 0 and 1. It may be noted that the change in opinion will be positive if the initial opinion is less than C, negative if the initial opinion is greater than C. In either case, however, the effect of the communication is to move the opinion closer to C.

To apply Equation 1, consider two communications, A and B, which may be given in AB or in BA order. Denote their fixed points by C_A and C_B, and their susceptibility coefficients by S_A and S_B, respectively. By renaming the communications, if necessary, we may assume that $C_B \geqq C_A$. This assumption is a matter of convenience, and entails no loss of generality. Its sole consequence is to give a positive rather than a negative sign to a recency effect. For any given initial opinion, denote the final opinion under the AB (BA) presentation order by X_{AB} (X_{BA}). The difference of these two final opinions is defined as the order effect. The algebraic expression for the order effect is

$$X_{AB} - X_{BA} = S_A S_B (C_B - C_A) \qquad [2]$$

Since all three factors on the right side of this equation are non-negative, the order effect will be nonnegative, so that $X_{AB} \geqq X_{BA}$. The magnitude of the order effect will increase with increasing S values, and with increasing distance between the two communications. It should be noted that the order effect does not depend on the initial opinion, except indirectly insofar as X_0 is correlated with susceptibility.

Equation 2 is derived by successive applications of Equation 1 as shown in Anderson and Hovland (1957). The reader may find it helpful to work through the following numerical illustration in which we take $C_A = 0$, $C_B = 10$, $S_A = \frac{1}{2}$, $S_B = \frac{1}{5}$, and an initial opinion equal to 5.

[1] This research was conducted during 1956–57 while the author was a Social Science Research Council Postdoctoral Training Fellow. Thanks are due C. I. Hovland for his many helpful ideas, and to Jack Brehm for his critical comments.

[2] Now at the University of California, Los Angeles.

Considering first the AB presentation order, the opinion after presentation of A will be

$$5 + \tfrac{1}{2}(0 - 5) = 2.5$$

The opinion, X_{AB}, after presentation of both A and B will then be

$$2.5 + \tfrac{1}{5}(10 - 2.5) = 4$$

Similarly, for the BA order, the opinion after presentation of B is

$$5 + \tfrac{1}{5}(10 - 5) = 6$$

Hence X_{BA}, the opinion after presentation of both B and A will be

$$6 + \tfrac{1}{2}(0 - 6) = 3$$

The order effect is thus

$$X_{AB} - X_{BA} = 4 - 3 = 1$$

It will be seen that this is, of course, the value obtained directly from Equation 2 when the given parameter values are substituted in it.

To apply Equation 2 to the primacy-recency problem, we suppose that A and B are communications on opposite sides of an issue. The assumption, $C_B \geqq C_A$, then simply amounts to orienting the opinion scale in such a way that B advocates an algebraically higher opinion than does A. By Equation 2, $X_{AB} \geqq X_{BA}$. Hence the communicator of B will prefer to act second since the final opinion, which will then be X_{AB}, will be greater and so more favorable to his position. Conversely, since $X_{BA} \leqq X_{AB}$, the communicator of A will also prefer to act second since the final opinion will then be X_{BA}, which is nearer to his position than is X_{AB}.[3] Thus, whenever Equation 1 holds, a recency effect will always be obtained.

Equation 2 may be applied as well to the weak-strong problem. In this case, it is supposed that A and B both advocate algebraically higher positions on the opinion scale, but with A weak and B strong, persuasibility-wise. It may be expected that the differences in strength of the two communications will be reflected in part in differences in their fixed points. Since X_{AB} is the greater of the two possible final opinions, it is seen that the weak-strong presentation order is the more effective.[4] In this case, the order effect would probably be relatively small since the difference between the fixed points, $C_B - C_A$, would not be expected to be as large as in pro-con experiments.

Formally, Equation 1 is identical with the linear operator equation of Bush & Mosteller (1955). It follows, in particular, that if a sequence of communications with the same S and C values were given, a common exponential growth curve for opinion as a function of "trials" would be obtained.

One final deduction from the model became important in the interpretation of the results, although it was not developed beforehand. Suppose one group of Ss is given the AB presentation order, another group is gven the BA presentation order. If each group is then administered the same sequence of further communications, the mean opinions of the two groups will converge toward equality. Because of the successive bifurcation procedure used in the experimental design, this general result implies that an order effect developed at any stage in the present experiment will decay to zero in the trials following. From Equation 1, the decay rate can be shown to be approximately equal to the average S value of the communications involved. For the present data, the S values are such that any order effect should become negligible in a few trials.

Before proceeding to test these predictions, it is well to consider possible objections to the model. The first of these is that the law of opinion change may not be linear in the given opinion scale as is assumed in Equation 1. Fortunately, at least three of the four predictions, excepting only the remark concerning exponential growth curves, are quite insensitive to the exact form of the equation. For instance, if X_0 lies between the two communications, a recency effect will obtain if amount of change is *any* increasing function of the distance between the communication and the S's opinion.

A more serious objection is that there may be systematic changes in susceptibility over

[3] This argument involves the implicit assumption, commonly made, that the goal of each communicator is to produce opinions as extreme as possible. More generally, one may desire to produce an opinion of some particular nonextreme value, in which case these considerations may not apply.

[4] See Footnote 3.

successive communications. The exact derivation of the order effect for this may be found in Anderson and Hovland (1957). However, it should be intuitively clear that decreases in S values will decrease the order effect relative to the base-line expression of Equation 2. If this decrease is large enough, the order effect will become negative, i.e., a primacy, or strong-weak effect.

One is thus led to expect that recency and weak-strong effects will be fairly generally obtained. They cannot be guaranteed, of course, without further knowledge about possible changes in susceptibility, but this problem lies beyond the scope of the present model.

The experiment reported here was designed to provide good data, and to test these various aspects of the model. Specifically, it was expected that recency and weak-strong effects would be obtained. More generally, it was hoped that the results would yield information about the adequacy of the linearity assumption of Equation 1, and about changes in susceptibility over successive arguments.

METHOD

The communication materials were taken from the case of Thomas Hoag, who was indicted for making a bigamous marriage to Catherine Secor. The case had been used in a persuasion experiment by Weld and Roff (1938), and their results were of considerable assistance in planning the present experiment. Most of the material given in Wigmore's (1937) summary was used to construct 17 pieces of evidence, here called arguments.[5] They consisted of the indictment, the summarized testimony of 14 witnesses (somewhat reworded, if necessary, to bring all to an approximately equal length of 175 words), and two pieces of court procedure which brought the trial to a close. The testimony of Benjamin Coe, witness for the prosecution, gives the flavor of the trial:

> Benjamin Coe testified that he was one of the judges of the court of common pleas in the county of Rockland; that he well knew the prisoner at the bar; that he came to Rockland in the beginning of September, in the year 1800, and there passed by the name of Thomas Hoag; that there was a person with him who passed for his brother, but between those two persons there was no sort of resemblance; that the prisoner worked for witness about a month, during which time he ate daily at witness's table, and he, of course, saw him daily; that on the 25th day of December, 1800, witness married the prisoner to one Catherine Secor; that witness is confident of the time, because he recollected that on that very day one of his own children was christened; that during all the time the prisoner remained in Rockland county, witness saw him continually; he was therefore as much satisfied that the prisoner was Thomas Hoag as that he himself was Benjamin Coe.

While not quite as strong as this, the remaining prosecution arguments were equally definite, and the testimony of the defense witnesses appeared to have no less a persuasive effect.

The Ss were Yale undergraduates, who were asked by phone to participate in the experiment. About 90% of those contacted accepted, and of these about 90% came to the experimental session. The procedure was administered to the Ss in groups of 3 to 13 at a time. Nine Ss were eliminated for skipping pages in the booklet, confusing the guilt-innocence orientation of the judgment scale, etc.

At the beginning of the experimental session, a brief account of orienting facts, important dates, and names in the case were listed on the blackboard where they remained throughout the session. General instructions containing the following points were then given: (*a*) the experiment was on the process of opinion formation; (*b*) although the evidence from a jury trial was being used, the Ss should not act as jurors, trying to withhold their judgment until all the evidence was in, but instead to give, each time they were asked for a judgment, their current opinion in the light of the evidence up to that point; (*c*) a description of the rating scale; (*d*) the case did not turn on some small detail or inconsistency of fact or date; (*e*) the case was genuine, not a mystery story, so that there was no "right" answer; (*f*) the case had been used with Cornell law students, and there was no final opinion which had not been defended by some of those students (see Weld & Roff, 1938).

Each S then received a booklet containing the arguments and read a practice argument written in the same style as the case, but unrelated to it and innocuous in content. After answering the Ss' questions, they were told that the remaining arguments were to be read and judgments made in time with E's signals. Conversation, and turning back or forward in the booklet, were disallowed. A general discussion of perhaps five minutes was held upon conclusion of the experiment, after which Ss were thanked for their cooperation and requested not to divulge details of the testimony until the conclusion of the experiment (at a given date).

For about the last two-thirds of the experiment, Ss were also asked to indicate on their booklets whether the time per argument had been too long, just right, or too short. The 247 usable replies were distributed 41%, 53%, and 6%, respectively, in the three categories.

Except as noted for Experiment 2, each page of the booklet contained one argument and a judgment scale. Two minutes were allotted each page, except that the two concluding arguments were given one, and one-half, minutes, respectively, because of their brevity. Judgments were made during the last 10 seconds of the allotted time. The judgment scale contained 11 cate-

[5] The texts of the arguments, the details of the construction of the sequences of arguments, and the cell means and analysis of variance tables for Exp. 1 have been deposited with the American Documentation Institute. Order Document No. 6036, remitting $2.50 for microfilm or $1.95 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

gories, with abbreviated descriptions of the meaning of each category, ranging from "belief in innocence beyond all reasonable doubt" to "belief in guilt beyond all reasonable doubt." INNOCENCE was written below the leftmost category, GUILT below the rightmost category. The arguments of the 14 witnesses were labelled PROS. or DEF., according as they were unfavorable or favorable to the defendant.

Experiment 1

The experimental variable was the order of arrangement of the prosecution and defense arguments.

Following the indictment, half of the *S*s received two prosecution arguments, then two defense arguments (PD order), while the other half received two defense, then two prosecution arguments (DP order). The difference in opinion of these two groups after the last of these four arguments is called the Stage 1 order effect.

Within each of these two groups, half the *S*s next received two prosecution followed by two defense, or else two defense followed by two prosecution arguments. The difference in opinion of these two pairs of subgroups after the last of these four arguments gives the Stage 2 order effect, unconfounded with presentation order in Stage 1.

In the same way, each of the above four subgroups was further split on Arguments 9–12, and the four pairs of differences in mean opinion after these four arguments give the Stage 3 order effect.

Following Stage 3, the same four arguments were given in the same order to all *S*s. The first two were prosecution arguments, while the last two brought in a revelation of innocence and concluded the experiment.

To safeguard against results accidental to a particular sequence of arguments, a number of different sequences were employed. The arrangement details were somewhat complicated and are only indicated below.[6] A fixed set of six prosecution and six defense arguments was used in Stages 1–3. Within each of the above eight subgroups, six different sequences of particular arguments were used. These sequences were so chosen as to evaluate the order effect at each of three stages as listed above, to completely counterbalance arguments across ordinal position, and also to yield an assessment of the weak-strong vs. strong-weak presentation order. This was accomplished by random formation of three pairs of prosecution and three pairs of defense arguments. Each pair was used in both forward and reverse order so that altogether 12 pairs were so formed, and these pairs constituted the unit in the construction of the sequences. To illustrate the procedure, let P_1, P_2, D_1, and D_2 denote particular prosecution and defense arguments. If one *S* received $P_1P_2D_1D_2$, in that order, then another *S* received $D_1D_2P_1P_2$, thus yielding a primacy-recency comparison after the fourth argument. Also, a third *S* received $P_2P_1D_2D_1$, and a fourth *S* received $D_2D_1P_2P_1$, yielding a similar comparison. At the same time, these sequences allow an evaluation of the weak-strong effect by testing P_1P_2 against P_2P_1, and D_1D_2 against D_2D_1.

In Stage 1, where there were only two different presentation order subgroups, there were only 12 distinct sequences. Each of these bifurcated, in harmony with the splitting of these two subgroups, to yield 24 distinct sequences leading through Stage 2. A similar bifurcation gave 48 distinct sequences leading through Stage 3. The order effect comparison at the end of each stage is thus between *S*s treated alike prior to that stage.

A random assignment of booklets was made in each batch of *S*s, with the restriction that one replication of the $8 \times 6 = 48$ distinct sequences was complete before another was begun. Four replications were run so that $N = 192$.

Experiment 2

The initial argument (indictment) and the last four arguments were the same as in Experiment 1. The intervening 12 arguments were presented either as six prosecution arguments followed by six defense, or else as six defense followed by six prosecution arguments. Six permutations of the defense and of the prosecution arguments were constructed randomly according to a latin square, so that each argument appeared equally often in each position. A randomly chosen two of the defense permutations followed each of two randomly chosen prosecution permutations, and conversely, so that there were 24 different particular sequences of arguments.

In addition to evaluating an order effect after 12 arguments, the experiment compared three judgmental conditions: Condition 1: judgments were made after each argument; Condition 2: judgments were made after the indictment, and the sixth, twelfth, and final arguments; Condition 3: judgments were made after the indictment, and the twelfth, and the final arguments.

For Condition 1, the booklets were similar to those of Exp. 1. For Conditions 2 and 3, the rating scales were cut off the bottom of the sheets containing the arguments, and sheets containing only a judgment scale were interleaved appropriately. *S*s in Conditions 2 and 3 were run in separate sessions since the judgment schedule was different. There were 48 *S*s in each condition, 24 in PD order, 24 in DP order, so that $N = 144$.

Experiment 2 was begun after completion of the first replication of Experiment 1, at which time it became clear that a sufficient number of *S*s would become available. Thereafter the two experiments were run concurrently.

RESULTS

The numbers 0 through 10 were assigned to the judgment scale categories for purposes of analysis, with 0, 5, and 10, in particular, standing for "complete belief in innocence", "neutral", and "complete belief in guilt", respectively.

The main results of Exp. 1 are shown in Figs. 1, 2, and 3, which plot mean opinion as a function of successive judgments, or trials. Successive figures show the mean opinion for the PD (denoted by O) and DP (denoted by X)

[6] See Footnote 5.

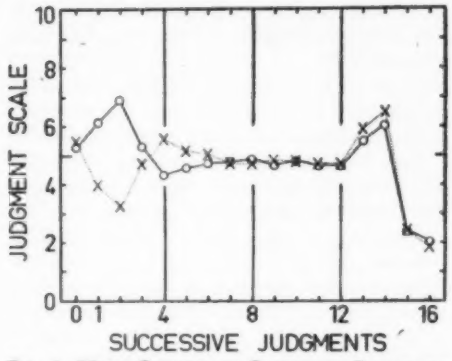

FIG. 1. MEAN OPINION ON SUCCESSIVE JUDGMENTS FOR THE TWO PRESENTATION ORDERS IN STAGE 1 OF EXP. 1; PD AND DP ORDERS DENOTED BY O AND X, RESPECTIVELY; LARGER NUMBERS ON THE VERTICAL SCALE INDICATE GREATER BELIEF IN GUILT

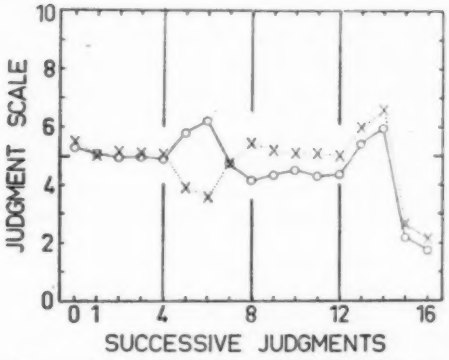

FIG. 2. MEAN OPINION ON SUCCESSIVE JUDGMENTS FOR THE TWO PRESENTATION ORDERS IN STAGE 2 OF EXP. 1
(Notation as in Fig. 1)

presentation orders within each of the three main stages of the experiment. In Fig. 1 are the results for the two presentation orders in Stage 1. Starting with a slight bias towards guilt, *Ss* in both conditions show marked changes. The PD *Ss* increase from an initial value of 5.31 to a value of 6.91 after two prosecution arguments, then decline to 4.36 under the influence of the two following defense arguments. The DP *Ss*, receiving the opposite order, swing in the reverse direction, and the crossing of the curves at J4 is the graphical equivalent of a recency effect. In the PD order, the defense, coming second, has the advantage

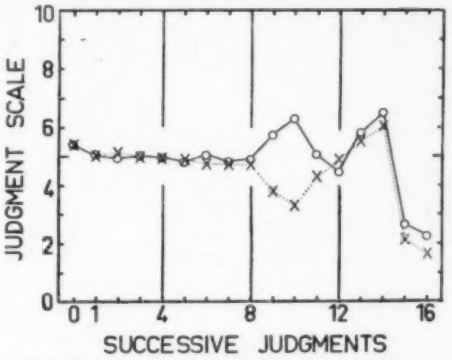

FIG. 3. MEAN OPINION ON SUCCESSIVE JUDGMENTS FOR THE TWO PRESENTATION ORDERS IN STAGE 3 OF EXP. 1
(Notation as in Fig. 1)

since the final opinion at J4 is more toward innocence. Similarly, the prosecution comes second in the DP order and there has the advantage, since in this case the opinion at J4 is higher, indicating a stronger belief in guilt. The difference of 1.22 in mean opinion for the two presentation orders at J4 gives the magnitude of the Stage 1 recency effect, which is significantly different from zero; see Table 1, Line A.

Following J4, the curves for these same two subgroups of *Ss* are averaged over their successive bifurcations in Stages 2 and 3. It is seen that the Stage 1 recency effect decays away in a few trials, and thereafter the two subgroups remain sensibly equal except perhaps at J13 and J14.

The Stage 2 results are given in Fig. 2. It will be recalled that there are actually four subgroups in this stage, corresponding to the splitting of each of the two Stage 1 subgroups at J4. Those two subgroups receiving the PD order in Stage 2 have been combined, as have those two subgroups receiving the DP order, and this same combining has been carried out in the preceding and following trials as well. The same pattern of behavior is found here as was obtained in Stage 1, although there is somewhat less yielding to the communications. Again a recency effect is found, of size 1.26, which is again significant. In contrast to Stage 1, this recency effect does not disappear in the succeeding trials, but after some initial decay,

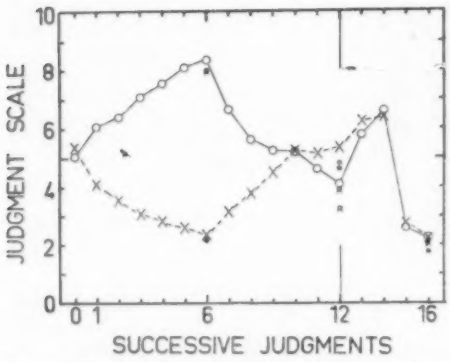

FIG. 4. MEAN OPINION ON SUCCESSIVE JUDGMENTS FOR EXP. 2. CONNECTED CURVES REPRESENT CONDITION 1, WITH PD AND DP PRESENTATION ORDERS DENOTED BY O AND X, RESPECTIVELY. SMALL FILLED CIRCLES REPRESENT CONDITION 2, SMALL UNFILLED CIRCLES REPRESENT CONDITION 3; EARED CIRCLES DENOTE PD ORDER, UNEARED CIRCLES DENOTE DP ORDER

is thereafter maintained without appreciable change till the last judgment.

Figure 3 compares similarly the two presentation orders in Stage 3, each curve being the average of the four subgroups which received the given presentation order in that stage. As before, the averaging was carried out over the prior and subsequent trials as well. Here also the order effect, measured by the mean difference at J12, is a recency effect. However, it is small (0.40) and falls short of significance. Differing in its further course from each of the two previous stages, the Stage 3 recency effect reverses to become a primacy effect in the four trials following. Thus, despite the fact that all *S*s were treated alike on these last four trials, the apparent advantage in being second in Stage 3 which appears at J12 is lost at once, and the final gain accrues to that side which was first in Stage 3.

Figure 4 displays the results of Exp. 2. The two order subgroups in Condition 1 first diverge, then converge and cross to yield a recency effect of size 1.21 at J12 which, however, disappears in the subsequent testimony. Condition 2, which is evidently no different from Condition 1 at J6, gives a similar recency effect at J12, and reaches essentially the same final level at J16 as does Condition 1. In Condition 3, the recency effect is again obtained at J12, and the final levels at J16 (omitted on the graph) are little different from the other four subgroups.

Table 1 gives the relevant quantities from the analyses of variance, and in each case includes all effects that were significant at the .05 level. Although the tests for the order effect itself in each stage are straightforward, the further breakdown of the treatment *SS* is more complicated and is not given in detail.

From Line A in Table 1, it is seen that the Stage 1 order effect is evidently real. The sequence effect indicates significant differences in the six means formed by pooling the paired PD and DP sequences which used the same particular arguments (e.g., $P_1P_2D_1D_2 + D_1D_2P_1P_2$). The nonsignificance of the interaction indicates that the magnitudes of the order effects for these same six pairs of sequences were not greatly different.

The analysis of the later stages is complicated in two ways. First, it would be expected that there will be a correlation between Stage 1 treatment and Stage 2 response, with *S*s who received the same arguments in Stage 1 responding more alike in Stage 2 than *S*s who received different Stage 1 arguments. For proper significance tests, a split-plot (Snedecor, 1956) analysis, of which the repeated measure-

TABLE 1
SUMMARY OF ANALYSES OF VARIANCE OF ORDER EFFECTS

Source	*df*	*F*	Error ms_e	Error *df*
A. Stage 1 (J6)				
Stage 1 order	1	24.58*	2.90	180
Sequences	5	6.54*	2.90	180
Interaction	5	1.65	2.90	180
B. Stage 2 (J8)				
Stage 2 order	1	23.83*	3.20	84
Sequences	5	6.44*	3.30	77
C. Stage 3 (J12)				
Stage 3 order	1	2.85	2.64	72
Stage 2 order	1	3.87	5.51	69
Sequences	5	2.21	5.51	69
D. J13–J16				
Stage 3 order	1	4.29*	38.55	72
Stage 1 order × Stage 3 order × Sequences	5	3.25*	38.55	72
Stage 2 order	1	4.93*	53.05	69
E. Exp. 2 (J12)				
Order	1	9.95*	3.92	72
Conditions	2	1.78	3.92	72
Conditions × Order	2	0.96	3.92	72

* $p < .05$.

ments design is a special case, is thus required. Second, there are 24 different sequences of arguments in Stage 2, so the over-all treatment *SS* has 23 *df*. Of these, 1 *df* corresponds to the mean Stage 2 order effect. The remaining 22 *df* may be further broken down into *SS*'s representing effects of Stage 1 order, sequences, and the interaction of these with Stage 2 order. Although this breakdown involves some confounding between Order in prior stages and Sequences, it increases sensitivity to possible inhomogeneities arising from the particular sequences of arguments used in the experiment.

The Stage 2 order effect is seen in Line B of Table 1 to be significant, as is the main effect of sequences. The near equality of the two error terms implies that the correlation mentioned above is small. Hence, the split-plot analysis on the J12 response in Line C allowed only for a possible correlation between *S*s treated alike in Stage 2, by considering as main plots the 24 different sequences administered in Stage 2. No significant effects were found at J12, although Stage 3 order, Stage 2 order, and sequences approach significance. The large difference in the two error terms indicates that *S*s treated alike in Stage 2 have a significantly correlated response at J12.

The analysis of Line D, carried out on the total score over the last four trials, shows that the Stage 2 order still has a significant effect, verifying, for these four trials, the observed retention of the Stage 2 recency effect seen in Fig. 2. The significance of Stage 3 order attests to the reality of the hidden primacy effect seen in Fig. 3. The significant interaction warns that these two effects may not be homogeneous over the sequences used, although the significance of 1 out of the 11 interactions evaluated in this analysis may be only a Type I error.

The retention of the Stage 2 order effect and the reversal of the Stage 3 order effect deserve further comment since they become important in the interpretation of the results. Although the Stage 2 recency effect falls slightly short of significance in the analysis at J12, this analysis uses only the data from that one trial. The effect is significant in the test over the last four trials because of the increased reliability of the score, even though the size of the effect is somewhat smaller, and the within-trial error variance is somewhat larger, than at J12. The Stage 3 primacy effect of the last four trials of Fig. 3 is the average of four comparisons, one from each of the paired subgroups in Stage 3. In the given analysis, all four comparisons were primacy effects. Indeed, making these four comparisons separately for each of the last four trials yields 16 comparisons. Of these, 14 are primacy effects, one is a recency effect, and one is zero. It is thus evident that the hidden primacy effect is not the result of one aberrant subgroup. Finally, it should be noted that this primacy effect is significant per se, over and above the J12 recency effect which it had to overcome.

As a check on the possibility that the *S* population became progressively contaminated by word of mouth report during the course of the experiment, the replication effect was tested in the analyses of Lines B, C, and D of Table 1, but the F was small and nonsignificant in each case.

The analysis for Experiment 2 is given in Line E of Table 1. No significant effects, except for sequences, were found at J6 or J16, so this information is omitted. At J12, the mean order effect is significant, but there were no significant differences between the order effects for the three conditions, or between their overall means.

Estimating the parameters, and testing the linearity assumption present considerable difficulty for a number of reasons. In particular, it should be noted that the procedure of assigning a priori values to the fixed points of the communications (such as $C = 0$ for defense arguments, $C = 10$ for prosecution arguments), which has been used by some workers, would be incorrect from the standpoint of the present model. However, some rough information in these two respects may be gained from inspection of Fig. 5. Here is plotted mean change in opinion as a function of initial opinion for each judgment in Stages 1–3 of Experiment 1, for each of the eight presentation order subgroups. In each panel, the points above the line represent the influence of prosecution arguments, those below the line correspond to defense arguments.

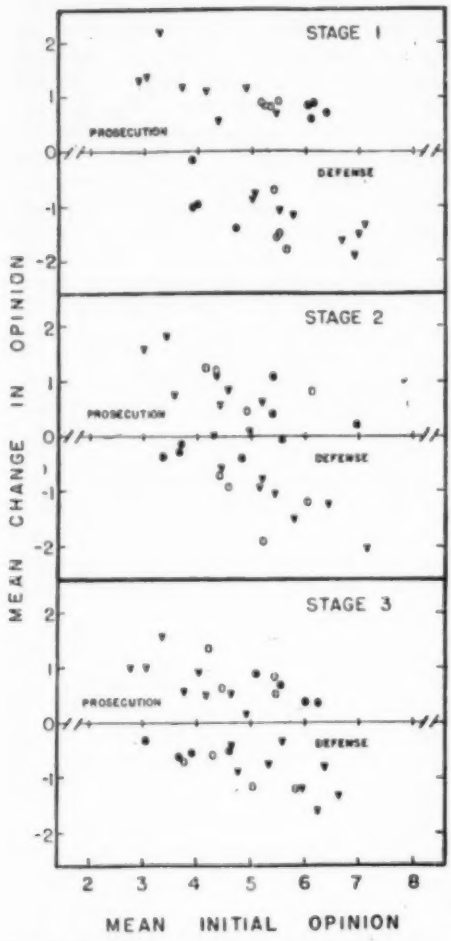

FIG. 5. MEAN CHANGE IN OPINION AS A FUNCTION OF MEAN INITIAL OPINION FOR EXP. 1. CIRCLE AND TRIANGLE DENOTE FIRST AND SECOND HALVES WITHIN EACH STAGE, RESPECTIVELY; OPEN AND FILLED FIGURES DENOTE FIRST AND SECOND ARGUMENTS WITHIN EACH HALF STAGE, RESPECTIVELY

The points are further classified according as they were in the first or second half of the given stage, and according as they were first or second within a half stage.

Ideally, were Equation 1 true, the six sets of points would lie on a straight line with slope equal to $-S$, and horizontal intercept equal to C. Fitting straight lines by eye suggest the following observations. The horizontal intercepts are somewhat closer to the center of the scale in Stages 2 and 3 than in Stage 1. Within each of the six sets of points, there appear to be systematic differences between the lines for the first and second half-stages (circle vs. triangle), with the line for the second half-stage showing intercepts closer to the center of the scale. Except for Stage 3, there do not seem to be marked differences between the lines through the points for the first and for the second arguments within a half-stage (open vs. filled symbols). Considered as a whole, no set or subset of points shows marked deviations from linearity.

The remarks of the previous paragraph suggest that the parameters are not constant over trials, and inspection of Figs. 1–3 suggests that there is a decreased over-all responsiveness in Stages 2 and 3 as compared to Stage 1. To test this, each S was assigned three total change scores by adding up his four opinion changes within each stage, giving to each a plus or minus sign according as it was or was not in the direction advocated by the argument. The mean total change scores for Stages 1, 2, and 3 were 4.40, 3.38, and 3.12, respectively. The difference between the first two means was significant, $F(1, 144) = 16.4$, while the F for the comparison of latter two was less than 1.

The interpretation of this decrease in total change scores as indicative of a decreased responsiveness needs justification, since there is involved a comparison of the amounts of change induced in opinions which are systematically different. The interpretation appears warranted here because of the pro–con counterbalancing in each stage, and because there is no marked asymmetry between the prosecution and defense arguments.

Since the three pairs of arguments of each type were given in both forward and reverse order, it is possible to test the relative effectiveness of the weak-strong and strong-weak presentation orders. Each argument was assigned a score equal to the total change which it induced over Stages 1–3. The members of each pair were classed as weak or strong according as this score was the smaller or larger. Members of one pair had identical scores and were omitted from the further analysis. Using

next the individual change scores induced by the pairs of arguments acting as a unit, the analysis of the mean over-all difference of weak-strong and strong-weak order yielded a significant $F(1, 720) = 6.71$, with the weak-strong order being the more effective. The actual effect was rather small, about 7% of the total change produced, as would be expected from the model.

DISCUSSION

As suggested by the model, recency effects were obtained at each of the three places at which it was planned to evaluate them. Support was also found for the predicted superiority of the weak-strong presentation order. To this extent, then, the predictive goodness of the model is substantiated.

The principal deviations from the model are seen in the further course of the order effects past the stages where they were formed. Although the decay of the Stage 1 recency effect is in accord with expectation, the maintenance of the Stage 2 recency effect throughout the remainder of the experiment, and the reversal of the Stage 3 recency effect in the subsequent testimony, are definitely contrary to the model.

The following, somewhat speculative, explanation of these discrepancies involves two assumptions one general and one specific. The general assumption states that opinion has two components: a basal component which is relatively little affected by the communications once it is formed, and a superficial component which is quite labile, and which obeys the original model. The two components act together to produce the observed opinion response.

Now at the beginning of the experiment, the basal opinion is presumably vague, or nonexistent, but begins to form as successive arguments accumulate. The special assumption locates the major portion of this basal opinion formation in the last half of Stage 2 and in the first half of Stage 3.

These two assumptions will account for the discrepancies according to the following argument. In Stage 1, there is as yet little basal opinion. Hence the obtained Stage 1 recency effect is created entirely by the superficial component. Since this component obeys the model, the decay of the Stage 1 recency effect is in accord with the prediction made in the introduction. The opinion at the end of Stage 2 has both superficial and basal components, and both contribute to the recency effect at this stage, the former because it obeys the model, and the latter because it is formed in the second half of the stage. In the succeeding arguments, that portion of the Stage 2 recency effect which is due to the superficial component decays away as seen in Fig. 2. The basal component, however, being resistant to change, is carried along beneath the perambulations of the observed opinion at a fairly constant value thereafter. The formation of the basal component in the first half of Stage 3 would of itself produce a primacy effect in this stage. The observed recency effect presumably arises from the overriding influences of the superficial component. Following Stage 3, the decay of the superficial component reveals the primacy effect caused by the basal component formed in Stage 3.

Although this explanation is *ad hoc*, there is an additional consideration which appears to lend support to the special assumption. On the basis of the linear model, it would be expected that the total change score and the order effect should be small or large together. It has been seen that the total change score is about equal in Stages 2 and 3, and in each smaller than in Stage 1. Yet, the recency effects are about equal in Stages 1 and 2, and in each considerably larger than in Stage 3. This relation of total change score and order effect would, however, seem to be consistent with the special assumption. For, if basal opinion is formed in the second half of Stage 2, it should act to enhance the recency effect relative to the average responsiveness in this stage. Also, if basal opinion formation occurs in the first half of Stage 3, it will act to reduce the recency effect relative to the responsiveness in this stage. This argument should be considered only as suggestive, however, since it cannot be made rigorous without a more precise specification of the way in which the two components jointly produce the observed opinion.

The central question for the model, as originally conceived, was whether the linearity

assumption would prove adequate, or whether it would be necessary to use a more complicated function to relate change in opinion to current opinion. The two-component hypothesis suggests a more basic problem, namely, that the structure of opinion may be too complex to be represented by a single number on an opinion scale. It should be noted that the idea of two independent components involves more than a simple increase in the "inertia" of the opinion. It had been realized that the opinion would probably become more resistant to change as its informational base accumulated over successive communications, but such an increased "inertia" could be represented in the model as a change in S values. One possible formulation would be to assume, as above, that Equation 1 holds for the superficial component, and that the basal component, although it comes into being, is largely resistant to change. Such an extended model would require provision for representing the amount of basal component present at any time, as well as the scale value of the opinion to which it corresponded. A rule for specifying how the two components combine to produce the observed response would also be needed. Under such a formulation, the present model would apply to situations in which the basal component either does not arise, or is already fully formed prior to the experiment.

Some comments on limitations and possible flaws in the results and interpretation deserve mention. It is in the first place evident that replication with other stimulus materials is needed. The present results may depend not only on having a situation in which there is essentially no initial opinion, but also on a set to reach a definite yes-no conclusion which might have been present. It is also possible that the results may arise specifically from the particular orders and sequences used. The general nonsignificance of the interactions in Experiment 1 gives some assurance as regards sequence peculiarities, but the nonappearance of a basal component in Experiment 2 suggests that the presentation orders used will be important. It should further be noted that, although the perseveration of the Stage 2 order effect virtually forces a two component assumption, these two components may correspond to two subsets of the S population rather than two opinion components within each S. Finally, even if the hypothesis of two components is basically correct, it may be oversimplified and in need of amplification.

With regard to the primacy-recency problem itself, these data make it amply clear that no general law of primacy or recency can exist, in agreement with Hovland's (1957) conclusion in his survey of the literature. Hovland goes on to suggest that recency effects will tend to be found when there is prior familiarity with the topic, and that primacy effects will tend to be obtained when the material is unfamiliar. The first suggestion would seem to be in accord with the model presented here as well as with the two component hypothesis since, with familiar material, there should be no great decrease in susceptibility and/or little formation of basal opinion consequent to the first communication. However, the second suggestion might well be weakened to state that with unfamiliar material, primacy effects are *possible*, but that the sign and size of the order effect will depend strongly on the particular arrangement of the sequence of communications, as well as on the time at which the opinion is measured. Here, with unfamiliar material, the particular arrangements used apparently acted so as to emphasize the recency effect in Stage 2, and to decrease the recency effect in Stage 3, with a subsequently appearing Stage 3 primacy effect. If the two-component hypothesis is reasonably correct, one would expect that primacy effects would be most easily obtainable at the locus of formation of the basal component.

In view of the above, it might be suggested that the order effect problem is not fundamental, but that order effects should be derived from more general principles. While this is probably true, the usefulness of the order effect paradigm as a tool in isolating such principles should not be overlooked.

Finally, it is worthwhile commenting on the finding in Experiment 2 that frequency of opinion measurement had no significant effect on the observed response, confirming and extending the result of Hovland and Mandell (1957). If this is further substantiated, it will allow a continuous record of the course of opinion formation without restricting the generality of the results.

SUMMARY

Two experiments designed to test a mathematical model for opinion change were presented. In each experiment, subjects read 17 successive arguments summarized from the testimony and procedure of a jury trial.

In Experiment 1, only presentation order was varied. In agreement with predictions from the model, recency and weak-strong effects were obtained at each place at which their evaluation had been planned. In further accord with the model, the recency effect formed in the first stage of the sequence of arguments rapidly decayed in the arguments following. However, the further course of the recency effects formed in the two later stages of the experiment was contrary to prediction. The recency effect from the second stage, after some initial decay, persevered at a fairly constant level through the remainder of the experiment. The recency effect formed in the third stage reversed itself to become a primacy effect in the trials following.

A two-component hypothesis about the structure of opinion was tentatively set forth to explain these discrepancies. According to this hypothesis, the opinion had a superficial component which obeyed the model, and a basal component, formed during the middle arguments of the experiment, which was largely resistant to change once it had come into being.

The order effect paradigm was used again in Experiment 2, which also compared three frequencies of opinion measurement. Recency effects were again obtained, but the frequency with which opinion was measured had no reliable effect.

REFERENCES

ANDERSON, N. H., & HOVLAND, C. I. The representation of order effects in communication research. In C. I. Hovland (Ed.), *The order of presentation in persuasion*. New Haven: Yale Univer. Press, 1957.

BUSH, R. R., & MOSTELLER, F. *Stochastic models for learning*. New York: Wiley, 1955.

FISHER, S., & LUBIN, A. Distance as a determinant of influence in a two-person serial interaction situation. *J. abnorm. soc. Psychol.*, 1958, **56**, 230–238.

FISHER, S., RUBINSTEIN, I., & FREEMAN, R. W. Intertrial effects of immediate self-committal in a continuous social influence situation. *J. abnorm. soc. Psychol.*, 1956, **52**, 200–207.

GOLDBERG, S. C. Three situational determinants of conformity to social norms. *J. abnorm. soc. Psychol.*, 1954, **49**, 325–329.

HOVLAND, C. I. Summary and implications. In C. I. Hovland (Ed.), *The order of presentation in persuasion*. New Haven: Yale Univer. Press, 1957.

HOVLAND, C. I., LUMSDAINE, A. A., & SHEFFIELD, F. D. *Experiments on mass communication*. Princeton: Princeton Univer. Press, 1949.

HOVLAND, C. I., & MANDELL, W. Is there a "law of primacy in persuasion"? In C. I. Hovland (Ed.), *The order of presentation in persuasion*. New Haven: Yale Univer. Press, 1957.

HOVLAND, C. I., & PRITZKER, H. A. Extent of opinion change as a function of amount of change advocated. *J. abnorm. soc. Psych.*, 1957, **54**, 257–261.

SNEDECOR, G. W. *Statistical methods*. (5th ed.) Ames: Iowa State Coll. Press, 1956.

WELD, H. P., & ROFF, M. A study in the formation of opinion based upon legal evidence. *Amer. J. Psychol.*, 1938, **51**, 609–628.

WIGMORE, J. H. *The science of judicial proof*. (3rd ed.) Boston: Little, Brown, 1937.

(Received September 2, 1958)

SOME EFFECTS OF INDIVIDUALLY PROMINENT BEHAVIOR UPON GROUP EFFECTIVENESS AND MEMBER SATISFACTION[1]

MARVIN E. SHAW

Massachusetts Institute of Technology

AN OBSERVER of individuals in small groups will notice a great variety of behaviors, so many in fact that the task of determining which behavioral characteristics are uniquely prominent in group interaction appears hopeless. Nevertheless, there is increasing evidence that the actual number of independent dimensions needed to account for the variance among behaviors is quite small. Carter (1954) reports the results of a factorial study which indicates that only three such dimensions are required; Individual Prominence, Group Goal Facilitation, and Group Sociability. A number of other studies (Clark, 1953; Hemphill & Coons, undated; Sakoda, 1952) support Carter's findings.

The results of factor analytic studies are of great importance in identifying the dimensions of behavior in small group interaction; however, they do not answer a question of crucial significance for the understanding of group behavior, namely: What are the consequences for group effectiveness and member satisfaction of varying degrees of behavior in each of the unique dimensions? This paper reports an attempt to determine some of the effects of individually prominent behavior upon problem solving efficiency and member satisfaction in small groups.

According to Carter (1954), Individual Prominence refers to a class of behaviors that are related to the person's efforts to stand out from others and to individually achieve various personal goals. In cooperative problem solving groups, this form of individually oriented behavior could be expected to interfere with effective group action. Thus it would be predicted that in unstructured[2] face-to-face groups, measures of group performance are negatively correlated with the degree to which group members indulge in individually prominent behavior.

The relationship between individually prominent behavior and member satisfaction is less clear; however, individuals who are prone to this form of behavior might enjoy a situation that affords the opportunity to satisfy this inclination, whereas individuals who are low with respect to this behavioral tendency might be unhappy with a group situation characterized by much individual-oriented behavior. If this reasoning is sound, we might expect a positive correlation between the tendency toward individually prominent behavior and satisfaction with the group situation.

The purposes of the experiment reported in this paper were (*a*) to test these expectations in small groups operating in an undifferentiated power structure, and (*b*) to determine whether the findings are different in groups operating in a power differentiated structure.

METHOD

The general plan of the experiment required four-person groups to solve problems in either a differentiated or an undifferentiated power structure; group performance and satisfaction scores were correlated with prior estimates of the degree to which group members tend to manifest individually prominent behavior. The tendency toward individually prominent behavior was estimated by means of a scale that is described below.

The IP Scale

The Likert procedure was followed in the development of the scale to measure Individual Prominence (IP). In selecting items, no attempt was made to conceal the purpose of the scale. It was believed that a person who wanted to mislead the test administrator by distorting his responses to the scale items would also distort his behavior in the experimental group situation. Since the purpose was to predict the person's behavior in the laboratory situation, it was felt that "obvious" items would discriminate at least as well as those whose meanings were hidden from the respondent. Also, originality was not mandatory; many of the items appear in other well-known personality and behavioral measuring devices, albeit in somewhat different form.

After preliminary investigation, 48 items were administered to 239 students at the Massachusetts Institute of Technology. An item analysis yielded 24 items which discriminated between the high and low scorers at the .01 level of significance. These 24 items make up the IP scale used in the study reported in this paper.

[1] This research was supported by a grant (G 2920) from the National Science Foundation.

[2] The term "unstructured" is used here to refer to the situation in which no structure has been imposed upon the group by the experimenter.

It contains items like: "In meetings of groups to which I belong I often find myself trying to get things organized." "In group situations I am quick to take the lead when an opportunity presents itself." "If I disagree with someone, I rarely let them know that I disagree." Both positive and negative items are included; agreement with the first two examples indicates high IP tendencies, whereas agreement with the third example indicates low IP tendencies.[3]

Split-half reliability was determined from the responses of a new sample of 270 students. It was found to be .79 when corrected for length by the Spearman-Brown formula.

Validity of the scale was estimated in two ways: (*a*) the scale was administered to three groups that were believed to differ with respect to Individually Prominent behavior and the group means were compared (concurrent validity), and (*b*) groups were observed and predictions based on scale scores were made about individual behavior (predictive validity). In the first test, mean scores of undergraduates, graduate students in management, and young executives were 76.7, 81.6, and 84.1, respectively. All differences were in the expected direction, but only the differences between the undergraduates and the other two groups attained acceptable levels of significance ($p < .01$).

In the second test, 20 four-person groups were observed in a situation which required that one member of each group volunteer for a task that would make him stand out from the other members. In addition to functioning as a group member, the volunteer was required to stand before the group and record on the blackboard the group's decisions regarding a number of issues. In 14 groups, the volunteer had the highest IP score in his group; in 4 groups, he had the second highest score; and in 2 groups, he had the third highest score. In no group did the person who had the lowest score on the IP scale volunteer for the recording job. This finding differs significantly from chance at better than the .001 level.

Although these tests do not demonstrate perfect validity, the validity of the IP scale appears to be satisfactory for research purposes.

Scores on the IP scale correlated +.02 with scores on the verbal scale of the Scholastic Aptitude Test ($N = 148$). Therefore, Individual Prominence probably is not related to intelligence.

Problems

The problems used in this experiment required *Ss* to make a series of decisions that determined the group score. They were seated before a blackboard upon which had been drawn a graph consisting of 100 three-inch squares. Lines along the abscissa were labeled with letters from A through J and the lines along the ordinate with numbers from 1 through 10. *E* had a similar chart drawn on a sheet of paper that *Ss* did not see. This chart contained a diagram consisting of four "concentric circles" similar to a bull's-eye target except that the circles were quite irregular and the bull's-eye might appear anywhere on the chart. Values were assigned to areas within each circle such that the coordinates outside the largest circle had a value of zero, those lying between the largest circle and the next one had the next lowest value, etc.; the bull's-eye had the highest value and was represented by a single coordinate. A different diagram was used for each problem.

On each trial, *Ss* chose a pair of coordinates, e.g., C5, D4, etc., the object being to earn as many points as possible. The same coordinates could be chosen as many times as desired; thus *Ss* could decide that the bull's-eye had been found and repeat his choice on all the remaining trials.

Subjects

The 176 *Ss* used in this experiment were male students at the Massachusetts Institute of Technology. Prior to the experiment they had taken the IP scale, but their scores were unknown to *E*. *Ss* were assigned to the various groups and conditions in haphazard order.

Procedure

When all four *Ss* in a group were seated before the blackboard, *E* explained the nature of the task and gave an example. When he was sure that everyone in the group understood the problem, he gave instructions designed to create the variations in power structure. Groups assigned to the undifferentiated power structure were instructed as follows:

> Each choice will be determined by group discussion. You will be given two minutes to discuss each choice. If you agree before the end of two minutes, we will go on to the next choice. If you do not agree by the end of the discussion period the choice will be the coordinates agreed upon by the largest number of persons. In case of a tie, I will toss a coin to decide which one to accept.

Groups assigned to the differentiated power structure were given the following instructions:

> Each choice will be made after group discussion. You will be given two minutes to discuss each choice, but if you finish before the end of the discussion period we will go on to the next choice. The actual decision regarding which coordinates to choose on any given trial will be made by ———. In making the decision he may consider the opinions expressed during the discussion period if he wishes; however, he will have complete and absolute authority to choose any coordinate he wishes, regardless of the opinions of other group members.

When *E* was satisfied that everyone understood the instructions, the discussion period for the first choice was begun. Each group was given 10 trials on each of five problems, requiring a total of 50 decisions. As each decision was made, *E* marked it on the blackboard and entered the score earned. At the end of each problem, *Ss* were permitted to examine the diagram which

[3] A copy of the complete scale has been deposited with the American Documentation Institute. Order Document No. 6038 from American Documentation Institute, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for 35-mm. microfilm or $1.25 for 6 x 8 photoprints. Make checks payable to Chief, Photoduplication Service, Library of Congress.

E had been using to determine the score for each decision.

At the close of the experiment, each *S* was asked to complete a questionnaire that required ratings of (*a*) personal satisfaction with the job, (*b*) group cooperation, and (*c*) group performance. *S*s were also asked to rank members of the group in order of amount of influence upon the group's decisions.

In addition to *E*, who gave the instructions and reported the score earned on each trial, an observer was present who sat at one side of the room and unobtrusively recorded selected aspects of the group interaction. Records were kept of suggestions made by each individual in the group, the order in which suggestions were made, the final decision on each trial, and the score earned. In addition, both *E* and *O* separately ranked *S*s in order of perceived influence on the group process.

RESULTS

The data were analyzed at both the individual and group levels. The purposes of the individual level of analysis were, first, to determine whether the IP scores were related to behavior in the group and, secondly, to determine whether satisfaction with the group situation as indicated by *S*'s ratings was related to the IP scores. Since the individual data were obtained from a number of subgroups, computation of the usual coefficient of correlation was not justified. Instead, a form of analysis suggested by Torgerson (1956) was used. The procedure is essentially a generalization of Kendall's tau to the case where the total sample is made up of several subgroups. This test does not yield a correlation coefficient, but it enables one to estimate the significance of the association between two variables.

The Torgerson test was applied to measures believed to reflect individually prominent behavior (number of suggestions, number of suggestions accepted by the group, and rankings of *S*s by observers[4] and by *S*s according to perceived influence), and to ratings of satisfaction, group cooperation, and group performance. Table 1 gives the probability that the true correlation between each of these measures and the IP score is zero. In the undifferentiated power structure, significant positive correlations were found between the IP score and all measures of individually prominent behavior. *S*s who scored high on the IP scale made more suggestions, had more of their suggestions accepted by the group, and were perceived by others as having more influence on the group's decisions. In the differentiated power structure, however, none of the relationships reached an acceptable level of significance. The imposed structure disrupted the tendency of the high IP *S*s to engage in behavior oriented toward individual prominence.

In the undifferentiated power structure, none of the ratings correlated significantly with IP scores. This finding clearly does not support the original hypothesis. In the differentiated power structure, IP scores were found to correlate significantly only with ratings of group cooperation. The reason for this significant correlation is not at all clear. It is possible that *S*s in the power structure were pleased that the "decision maker" allowed them to have some voice in the group decision and hence perceived the group members as being

[4] Rankings by *E* and *O* correlated .92 (average tau); therefore, the two rankings were combined for purposes of analysis.

TABLE 1

PROBABILITY THAT THE TRUE CORRELATION BETWEEN IP SCORE AND INDICATED MEASURE IS ZERO

Datum	No Power Structure	Power Structure
Number of suggestions per problem	.02 (+)	.37 (+)
Number of suggestions accepted by the group	.02 (+)	.85 (+)
Observer's ranking of *S*'s influence on group decisions	.02 (+)	.17 (+)
S's ranking of other *S*'s influence on group decisions	.01 (+)	.11 (+)
S's ratings of satisfaction	.12 (+)	.84 (−)
S's ratings of group cooperation	.80 (+)	.01 (+)
S's ratings of group performance	.65 (+)	.49 (+)

Note.—Symbols in parentheses indicate the direction of the correlation.

TABLE 2

CORRELATIONS BETWEEN MEAN IP SCORES FOR THE GROUP AND MEASURES OF GROUP PERFORMANCE

	No Power Structure	Power Structure
Percentage score	−.40*	−.07
Trials to target	+.41*	+.17

* Significant at better than the .05 level of confidence; the one-tail test was employed as direction of correlations was predicted.

cooperative. It is more likely, however, that this is just a "one-in-a-hundred" chance correlation.

At the group level of analysis, concern was directed to the relationship between the mean IP score for the group and measures of group performance. Two correlated measures of group performance were used: (*a*) the percentage score, which is computed as the ratio of points earned to total number of points possible, and (*b*) trials to target, which is defined as the number of trials required to locate the bull's-eye. Table 2 gives the correlations between mean IP scores and each of these measures in each of the two experimental conditions. In the undifferentiated power structure, mean IP was found to correlate negatively with the percentage score and positively with the trials to target score. These findings are in accord with expectations based upon the assumption that behavior directed toward individual goals interferes with effective group functioning. Imposition of the power structure negated these relationships.

DISCUSSION

Evidence from the individual analysis shows clearly that in the unstructured situation the higher the IP score the more likely the individual is to try to achieve individual prominence in the group. It is also clear that this relationship does not hold when one person in the group is given more decision making power than others. The person who is assigned to the decision making role takes a more active part in the interaction process than other group members, and hence tends to become more individually prominent regardless of his tendency in this direction (as indicated by his IP score). This relationship is shown by the fact that in the imposed power structure, the decision maker made more suggestions than anyone else in the group in 16 of the 22 groups. Since the chance expectation would be 5.5, this finding is highly significant ($\chi^2 = 26.6$, $p < .001$). On the other hand, the decision maker had the highest IP score in the group in only 5 of the 22 groups, a finding that is as near chance expectancy as possible. The greater activity on the part of a single member created by role assignment seems to account for the lack of relationship between IP score and individually prominent behavior in the power structure.

The lack of relationship between IP score and measures of satisfaction might be due to any of several reasons. One obvious reason is simply that the hypothesis was not valid. Another reason might be the insensitivity of the measures of satisfaction. The ratings of *S*s were remarkably homogeneous, more than 90 per cent being between 4 and 5. Since 5 was the maximum rating possible, the tests were based upon a very restricted range. Finally, it is possible that college students are more inclined than the general population toward individual prominence; if so, the range of IP scores is also restricted. Whatever the reason, however, the results of this experiment provide no evidence that the satisfaction of group members is related to the amount of individually prominent behavior in the group.

The data from this experiment provide good support for the hypothesis that the higher the mean IP of a group, the more poorly it performs. Much individually oriented behavior interferes with effective group action.

SUMMARY

An experiment was conducted to test the hypothesis that in unstructured situations, behavior oriented toward individual prominence in the group is negatively correlated with group performance and positively correlated with member satisfaction. A second purpose of the experiment was to determine whether similar relationships are found in groups differentiated along the power dimension.

Twenty-two four-person groups were observed under each structural condition. Groups were required to solve five problems requiring a total of 50 group decisions. After completion of the problem, *S*s rated their satisfaction with the group, the degree to which group members cooperated, and group performance. The tendency toward individually prominent behavior was estimated by scores on the IP scale, which was developed to measure this behavioral predisposition.

In the unstructured situation, mean IP score for the group was found to correlate negatively with measures of group efficiency, whereas in the power differentiated structure

no such correlation was found. Ratings of member satisfaction were unrelated to IP scores in both experimental conditions.

It was concluded that in the undifferentiated situation behavior oriented toward individual prominence interferes with effective group action; whereas, in the power differentiated situation, the effect of individually prominent behavior is reduced by the fact that the person who is given the decision making role becomes more individually prominent than others in the group, regardless of his usual tendencies in this direction. There was no evidence that individual prominence is related to member satisfaction.

REFERENCES

CARTER, L. F. Recording and evaluating the performance of individuals as members of small groups. *Personnel Psychol.*, 1954, **7**, 477–484.

CLARK, R. A. Analyzing the group structure of combat rifle squads. *Amer. Psychologist*, 1953, **8**, 333.

HEMPHILL, J., & COONS, A. *Leader behavior description*. Columbus: Personnel Res. Bd, Ohio State Univer. (undated).

SAKODA, J. M. Factor analysis of OSS situational tests. *J. abnorm. soc. Psychol.*, 1952, **47**, 843–852.

TORGERSON, W. S. A non-parametric test of correlation using rank order within subgroups. *Psychometrika*, 1956, **21**, 145–152.

(Received August 13, 1958)

MORPHEME ORDER AND SYLLABLE STRUCTURE IN THE LEARNING OF MINIATURE LINGUISTIC SYSTEMS

ARNOLD E. HOROWITZ AND H. MERRILL JACKSON

Mental Health Research Institute, University of Michigan

In 1925, Esper reported an experiment on the learning of Miniature Linguistic Systems (MLSs). An MLS is a set of nonsense syllables used as verbal symbols for a set of nonsense figures. Relations among the syllables, and between syllables and referents, are arranged in a manner similar to the structural features of natural languages and taught to subjects (*S*s) as the names of the figures.

In Esper's study, the first group was taught a system giving two-syllable names to 16 figures that were combinations of four shapes and four colors. The phonemes of the MLS all occur as phonemes in English, and the syllable structure followed the pattern of English. Esper interpreted the order of morphemes as also following English structure; the shape morpheme, equivalent to the English noun, followed the color morpheme, equivalent to the English adjective. The MLS in his second group gave two-syllable names to the same figures. However, some of the syllables and the morpheme order did not correspond to English structure. The contrast in morpheme order is exemplified by the following: in English one says "red circle" or "blue triangle" (Esper's first group) not "circle red" or "triangle blue" (his second group). The contrast in syllable structure between the two MLSs was produced by including ones with such non-English structure as initial [ng] and [mb].

Esper's results showed much faster learning for the first MLS. Also, in both groups, the shape morphemes were learned faster than the color morphemes.

Esper's design leaves open the question of exactly what was responsible for his observed differences. The non-English syllable structure *or* the non-English morpheme order (or both in conjunction) in the second MLS might have determined the longer learning time. There are several other possible sources of variation that could be controlled. According to the phonetic symbolism hypothesis (Sapir, 1949), there might have been some inherent relation between sound and meaning in the first MLS not present in the second. Such a relation would favor quicker learning. Heidbreder's work on concept attainment (1948) would imply that shape morphemes would be learned faster than color morphemes, but in Esper's experiment there was no control on the possibility that the syllables might have differed in ease of learning regardless of the referent dimension symbolized.

The present experiments were designed to isolate the MLS determinants of learning rate.

Experiment A[1]

Design

Four MLSs were set up and taught to separate groups of 10 *S*s each. The MLSs gave two-syllable names to nonsense figures that were combinations of five shapes and five colors. In these MLSs, *all* the syllables were structurally English, i.e., all phonemes were English sound units, and they were combined into consonant-vowel-consonant (/CVC/)[2] syllables according to a rigorous analysis of modern English structure (Trager & Smith, 1951). In two groups (I, Ia), the morpheme order was shape-color, as in Esper's slower group. In the other two MLSs (Groups II, IIa), the reverse morpheme order, color-shape, was used.

Two sets of five syllables ([A], [B] refer to the five members of each set as a group) were used as morphemes. In Groups I and II, the

[1] This experiment is part of a dissertation submitted by the senior author in partial fulfillment of the requirements for the Ph.D. degree in the Department of Social Relations at Harvard University. He is greatly indebted to the members of the department, particularly Richard L. Solomon, for guidance and support in carrying out the research. Financial support was received from funds of the Ford Foundation.

[2] Linguistic notational conventions use / / enclosing phonetic symbols to indicate reference to the phonemic level of a language; | | to indicate reference to the morphemic level. Reference to the phonetic character of sound units is indicated by square brackets.

TABLE 1
DESIGN FOR TESTING EFFECT OF SYLLABLE SETS AND REFERENT DIMENSIONS

Referent dimension	Syllable set	
	[A]	[B]
Shape	Groups I, IIa	Groups Ia, II
Color	Groups Ia, II	Groups I, IIa

position of syllable sets was [A # B],[3] and in Groups Ia and IIa it was [B # A]. These variations generate a four-fold table in which between-column difference in learning time would be attributable to the varying position of syllable sets, and between-row difference, to the variation in morpheme order.

Analysis of the learning times of separate morpheme types allows for isolation of two other possible determinants of Esper's results. Table 1 shows the design for this analysis. Between-column difference would indicate that syllable sets [A] and [B] differ in ease of learning regardless of the referent dimension they symbolized. Between-row difference would be due to difference in ease of learning the two referent dimensions, regardless of the syllables assigned as verbal symbols.

Phonetic symbolism is controlled both by random assignment of particular syllables to particular values of referent dimensions and by the fact that both sets of syllables were used as shape and color morphemes.

Thus one variable in Esper's experiment, English vs. non-English syllabification, was held constant in all these MLSs.

Materials and Procedure

The 10 syllables used in this experiment were chosen from 811 English nonsense syllables tested in a simple word-association procedure with 40 *S*s not used in the experiment proper. The syllables with the shortest reaction times were eliminated from use in our MLSs.

The nonsense figures were also pretested with the same 40 *S*s. The five shapes used were those with the least commonality of responses to the question, "What does this figure look like?" Each of the five shapes was made up in five colors: red, yellow, green, blue, and orange (Milton-Bradley Tonal Poster Papers 19P, 13P, 11P, 25P, and 15P). The shapes were traced on the poster paper, cut out, and pasted on five-by-seven inch white, nonglossy Bristol board cards. The cards were coated with clear acrylic spray. All figures were 7.85 sq. cm. in area.

E and *S* sat across a small library table. *E* presented referents by dropping the cards onto a study stand set up in front of *S*. The Learning Series consisted of 20 of the 25 combinations of shapes and colors. The other five figures were only shown to *S* after he had reached criterion in order to test for transfer of the learned system to new figures. There were 25 randomized orders of these 20 figures. These 25 orders were themselves given to *S* in a randomized sequence. Each time through the 20 figures is called a series, and series are the time units in our learning curves. During the first series, *E* presented the figures and pronounced the correct names, which *S* repeated. On every series after the first, *E* presented the figures and *S* had 15 seconds in which to give a name. At the end of the 15 seconds, or immediately after *S* gave a name, *E* gave the correct name and *S* again repeated it. The criterion was three series in which all 20 figures were correctly named. Shape- or color-morpheme types were considered to have reached criterion when all 20 of either type were correctly given in three series, regardless of whether the other morpheme type, or the complete names were also correct. *S*'s responses were recorded in phonetic notation. The measure of learning was the number of correct linguistic units per series. Analysis was made both of complete names and of each morpheme type separately, as required by our design.

*S*s were Harvard and Radcliffe students. Ages ranged from 17 to 36, with a mean of 21.3. No *S* had any speech impediment, color blindness, hearing defect, or uncorrected sight impairment that *E* could detect. All *S*s were native speakers of English. Differences in dialect and length and variety of second-language learning had no detectable effect on our results. There were five males and five females in each group. *S*s were paid one dollar per experimental session.

The Miniature linguistic systems.[4]

Syllable Sets: /A/; foyg, gef, ðɔp, sawb, dus
/B/; θuwž, kayv, šiyg, zɔwc, jahð

Morphology: Morpheme types and exhaustive lists are in Table 2.

Morphological structure. All names were of the form /CVC # CVC/.

Group I; {Shape} # {Color} using /A/#/B/.
Group Ia; {Shape} # {Color} using /B/#/A/.
Group II; {Color} # {Shape} using /A/#/B/.
Group IIa; {Color} # {Shape} using /B/#/A/.

Sample Names:

Referent; Shape A, red,
Group I /foyg/#/θuwž/
Group Ia /θuwž/#/foyg/
Group II /foyg/#/θuwž/
Group IIa /θuwž/#/foyg/.

[3] '#' is the notation used for a juncture between syllables; this is equivalent in length to the phrase pau se in English.

[4] Phonemic notation follows that of Trager and Smith (1951) though our interpretation of phoneme boundaries was somewhat broader than that in their microphonemic approach.

TABLE 2
MORPHEME TYPES AND LISTS

Type	Group I	Group Ia	Group II	Group IIa
Shape				
A	foyg	θuwł	θuwł	foyg
B	gef	kayv	kayv	gef
C	dəp	šiyg	šiyg	dəp
D	sawb	zɔwc	zɔwc	sawb
E	dus	jahd	jahd	dus
Color				
red	θuwł	foyg	foyg	θuwł
yellow	kayv	gef	gef	kayv
green	šiyg	dəp	dəp	šiyg
blue	zɔwc	sawb	sawb	zɔwc
orange	jahd	dus	dus	jahd

TABLE 3
MEAN NUMBER OF SERIES TO CRITERION

Group	Complete names	Color morphemes	Shape morphemes
I	10.1	9.8	9.4
Ia	13.6	11.5	12.2
II	10.2	9.4	9.0
IIa	11.8	10.2	10.0

TABLE 4
ANALYSIS OF VARIANCE: COMPLETE NAMES, NUMBER OF SERIES TO CRITERION

Source	*df*	*MS*	*F*
Position of syllable sets	1	66.77	2.907
Morpheme order	1	6.47	less than 1
Interaction	1	7.28	less than 1
Pooled error and interaction	37	22.97	

TABLE 5
MEAN NUMBER OF SERIES TO CRITERION FOR MORPHEMES GROUPED BY SYLLABLE AND REFERENT DIMENSION

Referent Dimensions	Syllables [A]	Syllables [B]	Both sets
Shape	9.7	10.6	10.15
Color	10.5	10.0	10.23
Both	10.75	10.00	10.06

Note.—$N = 80$, two morphemes per S.

Results

Table 3 presents the mean number of series to criterion for complete names and morpheme types separately. Analysis of variance applied to the number of series to criterion for the complete names yields the summary in Table 4. None of the Fs reached the .05 level. No effect can be attributed to either the position of syllable sets or the morpheme order variable.

The number of series to criterion for separate morpheme types, pooled according to Table 1, is shown in Table 5.

A summary of an analysis of variance is presented in Table 6. Again none of the Fs reached the .05 level and no difference can be attributed to either syllable sets or referent dimensions.

TABLE 6
ANALYSIS OF VARIANCE: MORPHEMES GROUPED BY SYLLABLE SET AND REFERENT DIMENSION

Source	*df*	*MS*	*F*
Syllable sets	1	1.02	less than 1
Referent dimensions	1	0.12	less than 1
Interaction	1	60.45	3.704
Pooled error and interaction	77	16.88	

The learning curves were Vincentized according to Hilgard's recommendations (1938). The first series, where S was first given the correct names, and the last series, where there was no possibility of variation, were not included in the computation. The remainder of an S's series was divided into fifths, the average number correct per fifth taken, and the mean of these individual averages taken as the point for the group. Curves for each morpheme type were computed on the basis of the complete name curve, even if they reached criterion first. This procedure gives the desired comparison between the parts in relation to learning the complete naming response. Figure 1 presents the Vincent curves for complete names and morpheme types. Mosteller[5] disagrees with Hilgard's dictum (1938) that statistics applied to Vincentized measures are invalid. In an attempt to get more information about possible differences in the learning process, we applied the method of allowances (Mosteller & Bush, 1954) to the Vincentized measures at each fifth of learning time. This analysis would show differences in rates of learning at different points of the curves. This sort of difference would not be reflected in the number of series to criterion measure. No differences among

[5] Mosteller, F. Personal communication. 1954.

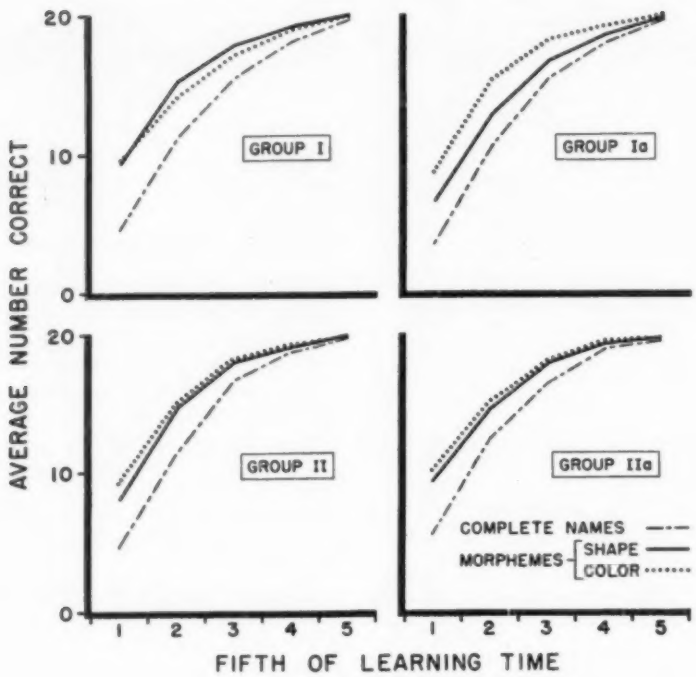

FIG. 1. VINCENT CURVES

the groups for either complete names or morpheme types were shown to be significant by this statistic.

To test Esper's finding that shape morphemes were learned faster than color morphemes in both his groups, comparison had to be made between the curves for the two morpheme types within each group. Application of statistical tests is therefore invalid since the two morphemes are part of the same "naming" response. The method of allowances can, however, be used as an "index" of difference.[6] Such a method is an improvement over simple curve inspection when dealing with such intractable data. It takes the range of scores into account as well as the group mean. The statistical computation was carried out on individual averages per fifth of learning time for the two morpheme types. The prior decision was made that unless the .05 level of differences was shown, the curves were not to be considered as being different at the given fifth of learning time.

[6] See Footnote 5.

Application of the above method to the two morpheme types within each group revealed only three points that differed according to this index. All three were in Group Ia. The level of learning of the color morphemes was higher than that of shape morphemes at Points 2, 3, and 5 of the Vincent curves. Except for this group, the differences between shape and color curves are extremely small, with no consistent trend in direction over the four groups. We can conclude, therefore, that there was no important difference in the rate of learning of the two morpheme types over the four groups.

Another check on a within-group difference in learning the two morpheme types was made. *S*s occasionally learned one or another of the morpheme types to criterion before reaching criterion for the whole naming response. If the same factor were at work in this experiment as caused Esper's difference in learning rate between shape- and color-morphemes, then it would be shown by our *S*s reaching criterion with shape morphemes before they reached it with color morphemes.

The differences with which morpheme types reached criterion were small, a matter, in most cases where it occurred,[7] of one or two series. This difference could not be expected to be statistically significant. Ranking the *order* in which shape and color morphemes reached criterion would, however, show any consistent difference in learning the two morpheme types. Kendall's (1948) coefficient of concordance (W) was used as a statistical test of the agreement of *S*s on the order in which they reached criterion for the two morpheme types. Ws for our groups were as follows: Group I, .041; Group Ia, .256; Group II, .091; Group IIa, .010. None of these Ws reach the .05 level. There was no tendency for either morpheme type to be learned before the other.

Discussion

Two major and three minor possible causes for the difference in rate of learning in Esper's experiment were either varied or controlled in this experiment. No difference in learning rate was shown by our *S*s. It remained at this point to design an experiment to pin down the other possible cause of his differences—variation in syllable structure.

EXPERIMENT B[8]

Materials and Procedure

To test the effect of variation in syllable structure, a design was set up in which *S*s were taught MLSs with first syllables of non-English structure.

The syllable set [B′] was constructed by altering those in set [B] as follows: /θuwž/ was changed to /npuwž/; /kayv/ to /zgayv/; /šiyg/ to /vkiyg/; /ɔwc/ to /mbɔwc/; and /jahd/ to /lgahd/. Thus all initial consonants in [B] were changed to clusters, none of which occur initially in English.

Two changes from Esper's non-English syllable structure were made. First, single consonants that cannot occur initially in English, e.g., /ŋ/ as in so*ng*, were avoided. Secondly, clusters with an initial stop cluster, e.g., /gd/, were not used because such a sequence is extraordinarily difficult for English speakers to articulate; at least a small vowel (/ə/ or /i/) will almost inevitably occur between the consonants. These changes in Esper's material avoided an extreme range of difficulty in pronunciation for English speaking *S*s. They also made first syllables, in one sense, structurally consistent. All non-English syllables are different from English syllables in the same way; they have a continuant-stop initial consonant cluster not occurring in English.

Two MLSs were set up and each was taught to 10 *S*s. The two systems had the same syllable sets in the same positions: [B′ # A]. Position of syllable sets was not varied since Experiment A showed no significant influence of this variable. Group IIIa $\bar{E}$[9] had the shape-color morpheme order and Group IVa $\bar{\bar{E}}$ had the color-shape order to determine if morpheme order has an effect when initial syllables are of non-English structure. The MLSs for these groups are those resulting from the substitution of the syllables in set [B′], for those in [B], in Ia and IIa in Table 2.

These two groups, along with Groups Ia and IIa from Experiment A, composed the design for Experiment B. Syllable structure and morpheme order variables were separated so that their effects could be tested independently. Group Ia had English syllable structure and shape-color morpheme order; IIa, English syllable structure and color-shape order; IIIa $\bar{E}$, non-English syllable structure and shape-color order; IVa, $\bar{\bar{E}}$, non-English syllable structure and color-shape order.

Groups IIIa $\bar{E}$ and IVa $\bar{\bar{E}}$ were made up of University of Michigan students. Since the results from Groups Ia and IIa (run in Cambridge) were to be used in this experiment, Group Ia was replicated with Michigan students to test for equivalence of populations. No significant differences between Cambridge and Ann Arbor Groups Ia were found on any of the measures used in Experiment A.

The age range of the Michigan students (Groups IIIa $\bar{E}$ and IVa $\bar{\bar{E}}$) was 20 to 33 years, with a mean of 23.8 years. In other details, *S*s paralleled those in Experiment A. These non-English groups were run by the junior author.

Preliminary runs showed that some *S*s do not learn the non-English syllable structure at all. They would "fixate" pronunciations with minimal vowels before or after the initial consonants. This would continue with the evident and expressed feeling that they were mimicing *E* perfectly. A procedure different from that used in Experiment A was therefore adopted: *S*s were run for a maximum of 25 series, the longest taken by any *S* in Experiment A. The number of *S*s who reached criterion within 25 series (successes) were then compared with the number who did not (failures).

TABLE 7

SUCCESS AND FAILURES IN ENGLISH AND NON-ENGLISH SYLLABLE STRUCTURE MLSs

	Successes	Failures	Total
$\bar{E}$	20	0	20
$\bar{\bar{E}}$	9	11	20
	29	11	40

[7] There was *no* difference between morpheme types for 11 of the 40 *S*s.

[8] This experiment was supported by the Mental Health Research Institute with funds from the Michigan State Legislature.

[9] These experiments are part of an ongoing research program and numbering of groups is consistent with the filing system being used.

Results and Discussion

Frequencies of successes and failures for the test of syllable structure effect are shown in Table 7. The chi square test was used to evaluate the proposed relationship, and indicates that there are significantly (beyond the .01 level) more successes with English syllable structure than with non-English.

With 15 successes for the shape-color order, and 14 successes with the color-shape order, there is obviously no effect of the morpheme order variable.

Esper's linguistic material, procedure, and design have been modified in these experiments. Linguistic material and procedure were made simpler, more specifiable, and more consistent. The changed design allowed independent test of the two variables. The difference between English and non-English syllable structure was the only variable producing a significant difference in these experiments.[10] To the extent that the modifications do not destroy comparability entirely, the implication is clearly that Esper's results were due to his syllable structure variable and not to morpheme order.

SUMMARY

Two experiments were conducted to test the effects of several variables confounded or uncontrolled in Esper's 1925 study of the learning of Miniature Linguistic Systems. Modification of the original linguistic material and procedure was made for the sake of simplicity and consistency. Of the possible causes of Esper's results, phonetic symbolism, particular grouping of nonsense syllables, particular referent dimensions, and morpheme order produced no effects in these experiments. The only variable causing a significant difference was whether or not the Miniature Linguistic System learned contained syllables of non-English structure.

[10] A recent study by Brown and Hildum (1956) demonstrated the significance of syllable structure in the perception of syllables as well as pronunciation as in this study.

REFERENCES

BROWN, R. W., & HILDUM, D. C. Expectancy and the perception of syllables. *Language*, 1956, **32**, 411–419.

ESPER, E. A. A technique for the experimental investigation of associative interference in artificial linguistic material. *Lang. Monogr.*, 1925, No. 1.

HEIDBREDER, E., BENSLEY, M. L., & IVY, M. The attainment of concepts: IV. Regularities and levels. *J. Psychol.*, 1948, **26**, 193–216.

HILGARD, E. R. A summary and evaluation of alternative procedures for the construction of Vincent curves. *Psychol. Bull.*, 1938, **35**, 282–297.

KENDALL, M. G. *Rank correlation methods.* London: Charles Griffin, 1948.

MOSTELLER, F., & BUSH, R. R. Selected quantitative techniques. In G. Lindzey (Ed.), *Handbook of social psychology.* Vol. I. Cambridge, Mass.: Addison-Wesley, 1954.

SAPIR, E. A study of phonetic symbolism. In D. G. Mandlebaum (Ed.), *Selected writings of Edward Sapir.* Berkeley: Univer. California Press, 1949.

THUMB, A. Psychologische studien über die sprachlichen analogiebildungen. *Indogerman. Forsch.*, 1907, **22**, 1–55.

TRAGER, G. L., & SMITH, H. L., JR. An outline of English structure. *Studies in linguistics*, Occasional Papers No. 3, Norman, Oklahoma: 1951.

(Received September 8, 1958)

AN EMPIRICAL TEST OF STRUCTURAL BALANCE IN SOCIOMETRIC TRIADS

STEPHEN H. DAVOL

University of Rochester[1]

THE present study is an empirical test, in sociometric triads, of Heider's theory of balance as defined in graph-theoretic terms by Cartwright and Harary (1956). In an extensive review of Heider's theoretical conception of balance, Cartwright and Harary point out the similarity between Heider's P-O-X unit and the A-B-X model in Newcomb's theory of interpersonal communication. In both theories, a tendency towards balance (or for Newcomb, a strain towards symmetry) is hypothesized. For Heider (1946), if an imbalanced state exists, forces arise to change the state towards balance either through cognitive reorganization or through some action; if such a change is not possible, the state of imbalance continues to produce tensions that make the situation uncomfortable.

Heider's conception of balance has been tested empirically by Kogan and Tagiuri (1958) using a P-O-Q unit where the X of Heider's P-O-X model was replaced by a third person, Q. In small group situations, three positive and three negative sociometric choices of each member were obtained, along with his perception of the choices made by the other members of the group. Using baselines of chance expectancy and of the "objective reality" provided by the sociometric choices, Kogan and Tagiuri found evidence that P's perception of the O-Q relationship significantly exceeded both baselines in the direction of the balanced configurations hypothesized by Heider. Balance occurred not only in the cognitive configurations of P, but also in the objective situation in the group itself, since the balanced configurations of the sociometric baseline exceeded chance.

Cartwright and Harary, using concepts from the mathematical theory of linear graphs (Harary & Norman, 1953; König, 1950), have presented a formal definition of balance which they feel may be applied in any situation where the terms "point" and "signed line" can be meaningfully coordinated to empirical data and thus with relationships such as Kogan and Tagiuri used for their sociometric baseline.

For the case in which the direction of the relationships is considered, Cartwright and Harary employ the concept of a directed graph, or digraph, which consists of "a finite collection of points together with a prescribed subset of the set of all ordered pairs of distinct points. Each of these ordered pairs AB is called a line of the digraph." In a directed graph coordinated with sociometric data, a mutual choice between A and B is represented by the two directed lines AB and BA. A signed digraph, or s-digraph, is one in which some of the lines are considered positive and some negative (the negative lines usually being represented as dashed lines). Figure 1 shows the 3-point s-digraphs used in this study. A semicycle is defined as a collection of lines in an s-digraph obtained by taking one line from each pair of points in the order A, B, C, A; the sign of the semicycle is determined by the product of the signs of its lines. In s-digraph c in Figure 1, a semicycle of length three and positive sign is $\overrightarrow{AB}$, $\overleftarrow{BC}$, $\overrightarrow{CA}$; a semicycle of length two and positive sign is $\overrightarrow{BC}$, $\overrightarrow{CB}$. Cartwright and Harary define an s-digraph as balanced if *all* its semicycles are positive. It should be noted that for a *triad* to be completely balanced for length three, each side itself must be balanced, i.e., the two directed lines between any pair of points must have the same sign. If any side or sides of a triad is imbalanced within itself, then the *degree* of balance for length three is .50, i.e., the number of balanced semicycles, four, divided by the total number of semicycles of the triad, eight.

The balanced configurations in the present study were derived from Cartwright and Harary's formal definition of balance in an s-digraph. The main problem of the study was to predict the occurrence exceeding chance of balanced configurations found in a "natural" situation of small groups living together. The predictions of balance in the 3-point s-digraphs

[1] Now at the Veterans Administration Center, Bath, New York.

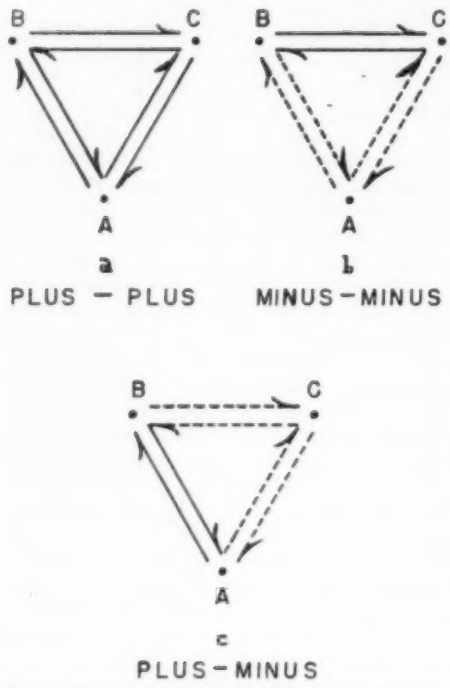

FIG. 1. THREE BALANCED S-DIGRAPHS OF THREE POINTS EACH DERIVED FROM THREE DIFFERENT AB, AC STRUCTURES

extracted from the small living groups were made as predictions of the two signs of the third side, BC, from knowledge of the signs of the other two sides, AB and AC. When the s-digraphs were restricted to those in which both the AB and AC relationships were symmetric, i.e., balanced for length two, predictions of balance could be made for three different AB, AC structures. In Fig. 1, the three 3-point s-digraphs are in balance for length three according to Cartwright and Harary's formal definition of balance in an s-diagraph. Three predictions are generated by these balanced s-digraphs: (*a*) when the relationships AB and AC are both mutually positive (the "plus-plus" s-digraph a of Fig. 1), the relationship BC will be mutually positive, (*b*) when the relationships AB and AC are both mutually negative (the "minus-minus" s-digraph b of Fig. 1), the relationship BC will be mutually positive, and (*c*) when the relationship AB is mutually positive and AC is mutually negative (the "plus-minus" s-digraph c of Fig. 1), the relationship BC will be mutually negative.

An additional problem was to assess the effect of the intensity or strength of the AB and AC relationships upon the occurrence of balance in the three predicted structures.

METHOD

Subjects and Procedure

The *S*s of the study were the veterans living in the rooms of a VA domiciliary building. The rooms varied in size from 7 to 20 men, with an average of 13. The men had lived in the rooms from 2 months to 10 years, had a median age of 63 years, and had a chronic disability which incapacitated them, preventing them from earning a living. The sociometric measure was given to the men in these living groups as part of the procedure of a larger study of the determinants of sociometric choice (Davol, 1957). The sociometric data in the present study of balance were from a selected sample of 141 of the *S*s.

The Syracuse Scales of Social Relations and Group Structure (SSSRGS) (Gardner & Thompson, 1956) were employed, in a form slightly modified for the domiciliary population. The situation on which the ratings were based[2] was very close to Gardner and Thompson's "affiliation" situation, and the reference population used by the *S*s in obtaining their own individual rating scale was: "All the men you have ever known." The SSSRGS yields a sociometric rating of each member in the group by each member in the group on an 8-point scale. A rating from 5 to 8 indicated acceptance, while a rating from 4 to 1 indicated rejection of a member of the living group for the "affiliation" situation given. For the purpose of determining balance in the s-digraphs resulting from the choices of the *S*s, ratings from 5 through 8 were considered directed lines of positive sign, and ratings of 4 through 1 were considered directed lines of negative sign. From the sociometric data, all mutual ratings of 8 were first selected; next, mutual ratings of 7 or 7 with 8; and finally mutual ratings of 6, 6 with 7, or 6 with 8. From these pools of mutual positive ratings at each intensity level, all the structures in which one *S* had two mutual positive relationships with two other members of his living group were extracted, i.e., where AB and AC were both mutually positive as in "plus-plus" s-digraph a of Fig. 1. The BC relationship for each occurrence of the predicted situation was classified into three categories, (*a*) mutually positive (any mutual rating of 5 or higher); (*b*) mutually negative (any mutual rating of 4 or lower); (*c*) unsymmetric or imbalanced for length two (a rating of 5 or higher with one of 4 or lower).

[2] There are some people with whom you like (or would like) to maintain a close and lasting friendship. You may not know why you like to be around these people, but you do know that being with them gives you an especially pleasant and satisfied feeling.

The procedure for selecting AB, AC structures was repeated for the 1 through 3 ratings, giving a pool of mutual negative relationships at each intensity level. From these pools of mutual negative relationships, all the structures where AB and AC were both negative (as in "minus-minus" s-digraph b of Fig. 1) were extracted, and the BC relationship for each occurrence of this structure was also classified into the same three categories, mutually positive, mutually negative, and unsymmetric.

Finally, from both the mutually positive and mutually negative pools of ratings, at equivalent levels of intensity, situations were extracted in which AB and AC were mutual relationships of opposite sign as in "plus-minus" s-digraph c of Fig. 1; and the BC relationship was classified as in the previous two situations.

RESULTS AND DISCUSSION

Since empirically there were an equal number of positive and negative ratings made by the *S*s taken as a whole, an equal number of mutually positive and mutually negative ratings could be expected to occur between B and C on a chance basis. Thus the tendency toward balance was tested at each level by the χ^2 test against the chance expectancy of an equal split of the total number of symmetric, or mutual, BC relationships at that level.

The analyses of the data indicate that only one of the three possible balanced triads (indicated in the predictions) actually tended toward balance in these "natural" situations. Table 1 shows the results for all three predicted structures at each intensity level, as well as at two accumulated levels. An inspection of the table, particularly the column of probability values for the χ^2 tests, shows that balanced structures did not occur uniformly in greater than chance frequencies for all levels of each predicted structure; therefore, any evaluation of the model's ability to predict must be made separately for each predicted structure.

TABLE 1

RESULTS OF GRAPH-THEORETICAL PREDICTIONS FROM THREE STRUCTURES AT THREE LEVELS OF INTENSITY

Level of Mutual AB and AC Relationships	BC Mutual Ratings				BC Unsymmetric Ratings	Total *N* of BC
	+	−	*N*	*P*[a]		
Prediction from "plus-plus" structure						
8	1	0	1		0	1
7	6	4	10	.70	8	18
7 & 8	7	4	11	.50	8	19
6	23	9	32	.02	23	55
6 & 7 & 8	30	13	43	.01	31	74
Prediction from "minus-minus" structure						
1	0	1	1		0	1
2	5	10	15	.20	4	19
1 & 2	5	11	16	.20	4	20
3	11	39	50	.001	35	85
1 & 2 & 3	16	50	66	.001	39	105
Prediction from "plus-minus" structure						
8–1	0	0	0		1	1
7–2	5	11	16	.20	18	34
8–1 & 7–2	5	11	16	.20	19	35
6–3	16	24	40	.30	47	87
8–1 & 7–2 & 6–3	21	35	56	.10	66	122

[a] Probability of deviation from chance expectancy using χ^2 test, 1 *df*.

Prediction from "Plus-plus" Structure

The graph-theoretical definition of balance is supported most strongly in the structure where AB and AC were mutually positive and a mutually positive relationship was predicted between B and C. The top section of Table 1 shows the prediction to be in the right direction at all intensity levels, but to be statistically significant only when the strength or intensity of the mutual relationships of AB and AC was at the level of six or when the results for all levels were accumulated. This somewhat surprising intensity effect may indicate the strength of the tendency towards balance in this type of sociometric structure compared with that of other forces responsible for the two mutually positive relationships. The results indicate that when the two positive relationships were strong, they were evidently more able to resist tensions brought about by the imbalance of a third mutual relationship of negative sign, but when AB and AC were weak, the tensions due to imbalance were of sufficient relative importance to bring about balanced triads in excess of chance. Considering the small absolute number of mutual relationships at level seven, the four imbalanced triads also could have been due to a competitive situation between B and C for the friendship of A. Whatever the source, forces were operating on the structure against the tendency towards balance and appeared to be operating more strongly where AB and AC represented mutually intense responses on the sociometric instrument. These forces need to be identified in future research and included in a more complex theory of balance.

Prediction from "Minus-minus" Structure

Taking next the s-digraph in which AB and AC are mutually negative relationships, the prediction of the BC relationship as mutually positive not only was unsupported by the data at all levels, but the opposite occurred significantly at the weaker level of intensity, three, as well as when all levels were taken together. This complete reversal of the graph-theoretical prediction is presented in the middle section of Table 1.

The predicted tendency toward balance in the "minus-minus" structure was not observed. Agreement on the rejection of another person, even combined with the joint rejection by that person, apparently is not sufficient grounds for creating a positive relationship between two people. Jordan (1953) and Hershkowitz (1954), in researches with Heider's theory, both found that a negative relationship is, in itself, uncomfortable. Starting with two mutual negative relationships in the structure, it is possible that the situation is sufficiently uncomfortable to make any strain towards balance irrelevant to the situation and that a tendency to reject the whole situation arises. In other words, withdrawing from the situation by rejecting both persons may be easier than trying to maintain a coalition against the third. The results, showing a greater imbalance at the weaker levels of the AB and AC relationships, suggest, however, that if the above explanation is the case, some tendencies toward balance may occur when strong negative feelings make agreement or coalition between B and C worthwhile.

Prediction from "Plus-minus" Structure

The results for the prediction in the "plus-minus" structure are shown in the bottom section of Table 1 and indicate that the prediction of a mutually negative relationship between B and C was in the right direction at both intensity levels, but only approached significance for the total accumulation of all levels. No increase in the proportion of relationships in the direction of the trend towards balance occurred when the AB and AC relationships in the s-digraph were weaker as had occurred in the other two structures. In fact, the intensity effect tended to work toward balance.

A discussion of the results for the third prediction should start with the fact that the final s-digraphs resulting from the second and third predictions are identical except for lettering. The difference between the two is the AB, AC structure from which the balanced triad is predicted. When the results of the predictions from the "minus-minus" structure are compared with those from the "plus-minus" structure, the tendency towards balance not only varies but is in the opposite direction for the two structures. Thus the data seemed to indicate that some positive relationship must be present for the tendency toward balance to operate. In the "plus-minus" structure, some positive feeling was already in the situation, and, in comparison to the "minus-minus" structure, a tendency towards balance appears in the data instead of the strong tendency towards imbalance. Apparently, once two members of the triad have some positive emotional investment in the situation, the forces toward balance tend to operate. Newcomb (1953) similarly has implied that the A-B-X model works best when some positive feeling is present in the situation.

The .50 Balanced Structure

In addition to the balanced and imbalanced s-digraphs found for each of the three predictions, there were .50 balanced s-digraphs which resulted from an unsymmetric relationship between B and C. The chance expectancy for the occurrence of this type of BC relationship was one-half of the total BC relationships found in the last column of Table 1. If imbalanced, or zero balanced, s-digraphs are expected to occur at less than chance frequencies and balanced, or 1.0 balanced, s-digraphs at more than chance, .50 balanced s-digraphs should occur at frequencies resembling chance expectancy. The last two columns of Table 1 reveal that this is very nearly the case. The data show only slight differences between the three structures in the occurrence of the unsymmetric BC relationships. The .50 balanced s-digraphs could be interpreted as structures in the process of moving toward balance or imbalance.

Generality of the Data

Before discussing the implications of the data taken as a whole, the generality of the results must be examined. The domiciliary living

groups are part of a definite subculture that has developed at the institution. Some preliminary data from a study in progress at the domiciliary indicate that the population has scores much higher on Srole's Anomie Scale (1956) than those found by Bell (1957) for a similar but uninstitutionalized population. The combined effect of these two factors, along with the advanced age of the men, seems to make for a lower level of interaction than would be expected in groups living together for such extended intervals of time. The subculture of of the domiciliary tends to support withdrawal as a means of dealing with interpersonal hostility and may have influenced the degree of imbalance found in the second and third predicted situations. This observation does not imply that structure is lacking within the domiciliary living units, as can be seen readily from the data of this study, but may mean that some of the choices are based on less behavioral interaction than would be expected in other living groups. However, the social structures that arise in domiciliary living groups certainly could be expected to follow in a general way the patterns found in other living groups where only moderate constraints are placed upon the group member to remain in the group.

General Implications of the Study

Although the groups used in the study lived in a particular subculture, the results have implications for the use of graph theory as a model for this type of data. The very fact that in these groups, only one of the three predictions was supported significantly by the data indicates that the context in which the relationships occur cannot be ignored in theoretical predictions of balance. Beyond this, the degree to which the results of this study have generality, the graph-theoretical model needs additional complexity to include the intensity of the relationships in the situation, and revision to predict more adequately situations in which negative relationships predominate. Cartwright and Harary (1956) mention "the development of a systematic treatment of relations of varying strength" as an unsolved conceptual problem regarding balance. The significant reversal of the prediction in the "minus-minus" structure casts serious doubt on the possibility of using graph theory as a source of predictions for the larger structures in these "natural" situations. Lack of balance at length three brings lack of balance to larger structures in which the three-point s-digraph is embedded and makes necessary accurate prediction at the lower levels before the more complex predictions at higher levels can be made. Furthermore, the results suggest that variables other than balance must be considered in predictions of sociometric structure. The data of this study indicate the need for studies having continuous measures of sociometric structures in the process of both growth and decay in order to clarify how balanced and imbalanced structures arise and change. Finally, the effect of balance at length three on the relationship between any two persons of a group can be considered as the summated effect of all possible three-point s-digraphs in which the two persons appear. Newcomb (1956) reported results supporting this position but also found contradictory results with strong negative relationships in his study of a college living group. No attempt was made to investigate the degree of balance in the summated effects in the present study.

Implications for the Theories of Newcomb and Heider

Since Newcomb's (1953) theory is one of interpersonal communication, the data have relevance to the theory and its strain towards symmetry to the degree that friendship choice is determined by communication between members of a triad. However, since Heider's (1946) theory of balance relates to the cognitive structure of a person, P, and thus his perception of the relationships in the situation rather than to reality, his theory is not affected by the results of this study. The results do have meaning for studies such as that of Kogan and Tagiuri (1958) where Heider's theory is tested against a sociometric baseline representing the actual balance in the situation. The results of this study demonstrate the necessity of looking for imbalance as well as balance and particularly the necessity for obtaining and reporting negative as well as positive relationships. The results also show that with traditional sociometric nominating techniques, a baseline is obtained which provides the most lenient test for Heider's perceptual balance.

Finally, the number of unsymmetric BC relationships that occurred with each of the three structures indicates that in determining balance, finding a positive or negative relationship between two persons in one direction is no guarantee that this relationship is reciprocal. There is no sure way to tell whether a choice is or is not reciprocated when sociometric nominations of three choices are used; and the incidence of unsymmetric relationships in this study indicates that the degree of balance found simply by counting the number of relationships with the predicted sign, as did Kogan and Tagiuri, will be inflated by the number of unsymmetric relationships in the system. Therefore, when sociometric data are used as a baseline for purposes of comparison with perceptual data, both intensity and reciprocation of choice must be considered in the interpretation of the data.

SUMMARY

Cartwright and Harary's graph-theoretical extension of Heider's theory of balance was tested in the three possible balanced sociometric triads. Three theoretical predictions were made from these triads. The empirical results showed that the "plus-plus" structure significantly tended toward balance, the "minus-minus" structure tended significantly toward imbalance, while the "plus-minus" structure tended toward balance but was not significant. The results indicated that unsolved problems exist regarding the intensity of relationships, the social context in which the relationships occur, and the effect of structures composed primarily of negative relationships. Experimental and theoretical implications of the findings are discussed, including the particular need for research on the growth and decay of sociometric structures.

REFERENCES

BELL, W. Anomie, social isolation, and the class structure. *Sociometry*, 1957, **20**, 105–116.

CARTWRIGHT, D., & HARARY, F. Structural balance: A generalization of Heider's theory. *Psychol. Rev.*, 1956, **63**, 277–293.

DAVOL, S. H. Some determinants of sociometric choice in a Veterans Administration Domiciliary. Unpublished doctoral dissertation, Univer. of Rochester, 1957.

GARDNER, E. F., & THOMPSON, G. G. *Social relations and morale in small groups*. New York: Appleton-Century-Crofts, 1956.

HARARY, F., & NORMAN, R. Z. *Graph theory as a mathematical model in social science*. Ann Arbor, Mich.: Inst. soc. res., 1953.

HEIDER, F. Attitudes and cognitive organization. *J. Psychol.*, 1946, **21**, 107–112.

HERSHKOWITZ, A. Interpersonal agreement and disagreement on objects and values: A preliminary study. Unpublished doctoral dissertation, Univer. of Kansas, 1954.

JORDAN, N. Behavioral forces that are a function of attitudes and cognitive organization. *Hum. Relat.*, 1953, **6**, 273–287.

KOGAN, N., & TAGIURI, R. Interpersonal preferences and cognitive organization. *J. abnorm. soc. Psychol.*, 1958, **56**, 113–116.

KÖNIG, D. *Theorie der endlichen und unendlichen Graphen*. Leipzig, Germany: Akademische Verlagsgesellschaft, M.B.H., 1936. (Republished: New York: Chelsea, 1950).

NEWCOMB, T. M. An approach to the study of communicative acts. *Psychol. Rev.*, 1953, **60**, 393–404.

NEWCOMB, T. M. Prediction of interpersonal attraction. *Amer. Psychologist*, 1956, **11**, 575–586.

SROLE, L. Social integration and certain corollaries: An exploratory study. *Amer. sociol. Rev.*, 1956, **21**, 709–716.

(Received September 17, 1958)

EXTRAVERSION, NEUROTICISM, AND MOTOR LEARNING

A. W. BENDIG AND CHARLES J. VAUGHAN
University of Pittsburgh

THE integrative attempt of Eysenck (1956) to relate the personality dimensions of extraversion and neuroticism to learning and perceptual phenomena provides a large number of experimentally testable deductions. Eysenck firmly associates extraversion with the rate of development, total magnitude attained, and slowness in dissipation of reactive inhibition in learning phenomena and less surely accepts the position that "manifest anxiety" or neuroticism is associated with total drive level. From the deduction that reminiscence phenomena in motor learning should be related to questionaire measures of extraversion and neuroticism, three studies have been reported (Eysenck, 1956; Jensen, 1958, pp. 299-300) that appear consistent with the hypotheses, but the relationships found are still too tenuous for complete confirmation of the asserted connections between temperament and learning variables.

This work has concentrated almost exclusively on reminiscence measures, or the difference between performance on the last trial of an initial period of motor practice and the first trial of a subsequent period of practice with a fixed rest (nonpractice) interval intervening between the practice periods. Little has been reported on the relationship between measures of extraversion and neuroticism and performance during the original period of motor practice on the task, although Eysenck noted nonsignificant negative correlations of −.10 and −.15 between questionnaire measures of extraversion and neuroticism and a practice score obtained by subtracting initial from terminal performance scores during the initial motor practice period (Eysenck, 1956, p. 331). The generation of appreciable amounts of reactive inhibition during the initial practice period does influence the magnitude of reminiscence effects (Albright, Borresen, & Marx, 1956; Archer, 1954; Kientzle, 1946; Kimble: 1949a, 1949b; Schucker, Stevens, & Ellis, 1953; Wasserman, 1951) but it also appears to be related to the acquisition of performance during the initial practice period in these same studies. It may be argued that reminiscence variables are more sensitive measures of the amount of reactive inhibition than are variables related to the original acquisition performance, but individual personality differences that are related to reminiscence measures should show similar, albeit weaker, relationships to acquisition measures. The psychometric reliability of reminiscence measures can also be questioned since they are scores computed as the difference between two relatively unreliable performance trials (Mosier, 1951). Indeed, Eysenck (1959, p. 330) found a correlation of only .44 between two successive reminiscence measures for his sample of 50 *S*s.

EXPERIMENT I

Method

Subjects. The 72 *S*s (45 men and 27 women) were undergraduate students in two sections of a one-semester introductory psychology course. All had been tested during the second week of the semester on the Maudsley Personality Inventory (Eysenck, 1956, 1957) of Extraversion and Neuroticism. These scales had moderately high reliability and a relatively low intercorrelation in other samples (Bendig, 1957). None of the *S*s were aware of their scores on the MPI at the time that the learning data were collected in the classroom approximately two months later.

Learning task. The inverted alphabet printing task devised by Ruch and Warren (1941) and utilized by others in studies of reminiscence phenomena (Albright, et al., 1956; Archer, 1954; Kintzle, 1946; Kimble: 1949a, 1949b; Ruch & Warren, 1941) was used to provide measures of motor learning. Each *S* was given a booklet containing 11 pages: (*a*) a face sheet containing the instructions, an example of the inverted first seven letters of the alphabet, and spaces for practicing one complete sequence of the inverted alphabet, and (*b*) 10 pages for recording responses during the learning trials with each page containing 11 rows of 11½-in. sq. *S*s were first asked to number the pages from 1 through 10, and *E* read the instructions aloud and answered any questions from *S*s. The instructions were those taken from other studies using this learning task (Kimble, 1949a; Wasserman, 1951). A single 60-sec. trial was given on each page with *E* timing each with a stop watch and calling out, "Stop! Turn to the next page", after each 1-min. trial.

Quantification. Two learning measures were derived

from *Ss*' performance. Learning Total was the total number of alphabet letters printed by each *S* during the 10-min. practice period. Learning Rate was the difference between the number of letters inverted and printed during the first two minutes of practice (Trials 1 and 2) and the last two minutes of practice (Trials 9 and 10). It was expected that Learning Total, based upon performance over all 10 trials, would be a more reliable measure than Learning Rate, which was computed from only four trials, but that Learning Rate would partially equate the *Ss* for initial rate of responding and provide a better measure of response acquisition over the practice period for individual *Ss*. Learning Rate is also a measure similar to Eysenck's (1957, p. 331) practice score in his study of pursuit-rotor performance.

Results

The two sex groups of *Ss* were first tested for significant differences in their means and variances on each of the measures. The women *Ss* ($N = 27$) had a significantly (.01 level) higher mean on the Extraversion scale ($t = 3.06$), and the men ($N = 45$) were significantly (.01 level) lower on Learning Total ($t = 3.56$). The sex differences in means on the Neuroticism scale ($t = .33$) and on Learning Rate ($t = .16$) were not significant. None of the F ratios comparing sex differences in the variances of the four test and learning measures approached significance at the .10 level. The highly significant sex differences in means on two of the four variables precluded the pooling of the *Ss* in subsequent analyses of the data.

The internal consistency of each of the four variables was separately assessed for the men and women. The reliabilities of the MPI scores were computed by the modified Kuder-Richardson Formula 20 (Bendig, 1957). The reliabilities of the Extraversion scale for the men and women *Ss* were .73 and .75, while coefficients for the Neuroticism scale were .86 and .79. The reliabilities of the Learning Total scores were computed by the usual analysis of variance procedure (Thorndike, 1949, pp. 93–96) and were .95 for each of the sex groups. Reliabilities for the Learning Rate variable were determined by correlating the scores on Trials 1 and 2 ($r = .62$ and .70), correlating Trials 9 and 10 ($r = .83$ and .81), correcting these coefficients by the simplified Spearman-Brown formula ($r = .76$, .82, .91, and .89), correlating the summed scores on Trials 1 and 2 with the summed scores on Trials 9 and 10 ($r = .47$ and .56), and inserting these coefficients in the formula (Mosier, 1951, p. 777) for the reliability of difference scores. The resulting Learning Rate reliabilities for the two sex groups were .69 and .68.

The intercorrelations among the measures were computed separately for the two sex groups. The intercorrelations of Extraversion and Neuroticism were −.25 and −.29, neither significant at the .05 level. The correlation between Learning Total and Learning Rate was .30 (significant at the .05 level) for men and .15 (not significant) for women. Correlations between the test variables and the learning measures were nonsignificant and uniformly low with the mean correlations (averaging the coefficients from the two sex groups) of Extraversion and Neuroticism with Learning Total being −.05 and .02 and the MPI correlations with Learning Rate being .09 and −.03.

The *Ss* within each sex group were dichotomized on both MPI variables, using the median scores found for a larger university norm group of *Ss* ($N = 1064$) to whom the MPI had been previously administered. The raw score ranges defining the high Extraversion groups were 30–48 for the men and 31–48 for the women Raw score ranges for the high Neuroticism groups were 20–48 (men) and 21-48 (women). The Learning Total and Learning Rate measures were analyzed as a $2 \times 2 \times 2$ (Extraversion, Neuroticism, Sex) factorial design, using a fixed effects model (Stanley, 1956) and the computational procedures discussed by Walker and Lev (1953, pp. 381–382) for analyses of variance with disproportionate subclass frequencies. The results of these analyses can be found in Table 1. The only significant source of variation for Learning Total was the main effect of sex differences (.001 level), which would be expected from the results of the t tests reported above. Neither of the main effects of Extraversion or Neuroticism nor any of the interactions approached statistical significance. In the analysis of the Learning Rate measures, only the interaction of Extraversion and Neuroticism approached significance, and this effect was significant only at the .10 level. To elucidate the reason for this inflated interaction term, the mean Learning Total and Learning Rate scores for each of the sex and MPI subgroups are given

TABLE 1
ANALYSES OF VARIANCE OF LEARNING MEASURES FOR EXPERIMENT I

Source of Variation	df	Learning Total		Learning Rate	
		MS	*F*	*MS*	*F*
Extraversion (E)	1	15.9	.02	7.4	.17
Neuroticism (N)	1	218.7	.24	.9	.02
Sex (s)	1	11162.4	12.46**	7.3	.17
E × N	1	16.5	.02	125.5	2.88*
E × S	1	108.9	.12	.7	.02
N × S	1	133.1	.15	25.6	.59
E × N × S	1	1145.0	1.28	20.1	.46
Within subgroups	64	895.5		43.6	

* $P < .10$.
** $P < .001$.

TABLE 2
MEAN LEARNING MEASURES OF MPI SUBGROUPS FOR EXPERIMENT I

MPI Subgroups		Learning Total			Learning Rate		
Extraversion	Neuroticism	Men	Women	Total	Men	Women	Total
High	High	365.3	479.5	411.0	25.7	33.8	28.9
High	Low	405.0	455.0	433.6	39.7	34.2	36.6
Low	High	390.9	442.5	403.8	38.1	41.0	38.8
Low	Low	388.5	471.6	412.9	29.9	32.0	30.5
High		389.1	461.1	426.3	34.1	34.1	34.1
Low		390.0	455.7	407.6	34.8	36.9	35.4
	High	384.5	457.3	405.9	35.0	38.1	35.9
	Low	395.6	459.9	424.3	34.1	33.6	33.9
Total		389.7	458.9	415.7	34.6	35.3	34.8

in Table 2. It can be seen that Learning Rate is low for the *S*s high on both MPI scales and also low for *S*s with low scores on both Extraversion and Neuroticism. This phenomenon is responsible for the sizable E × N interaction mean square in Table 1, but although this pattern of subgroup means is shown by both sex groups, this effect is not significant at the conventional .05 level.

Discussion

The above results only weakly support the hypothesis that measures of Extraversion and Neuroticism are related to motor learning performance. In the analyses of variance, neither of the main effects of Extraversion and Neuroticism showed any relationship to either Learning Total or Learning Rate, and the rectilinear correlations between the psychometric and learning variables were uniformly low. The only possible significant finding was in the interactive effect of Extraversion and Neuroticism upon Learning Rate. However, we would hesitate to suggest that this interactive effect is more than a sampling fluctuation because (*a*) it is statistically significant only at the .10 level, and (*b*) the interactive effect shown in the means in Table 2 cannot be deduced from Eysenckian theory. The finding that high scores on both MPI personality variables or low scores on both variables correlate with a slower rate of learning merits further study. In line with the possibility of a real interactive effect, highly similar patterns of subgroup means were found for both men and women.

The highly significant sex difference for Learning Total confirms the results reported by Archer and Bourne (1956) and is interesting in view of the mixed sex groups used by other investigators (Albright, et al. 1956; Archer, 1954; Kintzle, 1946; Kimble, 1949a, 1949b; Schucker, 1953; Wasserman, 1951) in studies of motor learning phenomena. If this sex difference on the inverted alphabet printing task is not a sampling error, then the intergroup differences reported in these studies may be attributable to differences in percentages of men and women *S*s in the groups.

EXPERIMENT II

The second study was designed as a repetition of Experiment I in order to examine two possibly significant findings in the first study: the sex difference in Learning Total and the interactive effect of Extraversion and Neuroticism upon Learning Rate.

Method

The learning task employed in Experiment I was used with 99 *S*s (62 men and 37 women) in three sections of introductory psychology. All of the details as to administration of the MPI and the inverted alphabet printing task described in Experiment I were duplicated for the present sample. Quantification of Learning Total and Learning Rate also was identical.

Results

The analyses of variance of the learning measures are given in Table 3. Again, there was

TABLE 3
ANALYSES OF VARIANCE OF LEARNING MEASURES FOR EXPERIMENT II

Source of Variation	df	Learning Total		Learning Rate	
		MS	F	MS	F
Extraversion (E)	1	1862.0	2.63	20.8	.72
Neuroticism (N)	1	14.5	.02	31.0	1.07
Sex (S)	1	2836.2	4.00*	113.9	3.93*
E × N	1	650.7	.92	1.1	.38
E × S	1	19.7	.03	31.0	1.07
N × S	1	55.4	.08	22.5	.78
E × N × S	1	217.0	.31	69.9	2.41
Within subgroups	91	708.3		29.0	

*$P < .05$.

TABLE 4
MEAN LEARNING MEASURES OF MPI SUBGROUPS FOR EXPERIMENT II

MPI Subgroups		Learning Total			Learning Rate		
Extraversion	Neuroticism	Men	Women	Total	Men	Women	Total
High	High	377.1	395.9	386.1	28.0	34.2	31.0
High	Low	382.1	432.3	404.4	25.9	26.9	26.3
Low	High	351.0	397.0	365.3	32.4	34.7	33.1
Low	Low	340.9	376.5	348.8	20.0	40.8	24.6
High		379.8	414.1	395.6	26.9	30.5	28.6
Low		346.9	390.7	359.0	27.3	36.5	29.9
	High	361.3	396.4	374.9	30.7	34.4	32.6
	Low	362.2	418.4	382.2	23.0	30.4	25.1
Total		361.7	405.9	378.2	27.1	32.6	29.2

a significant sex difference in Learning Total, but this time the F ratio was significant only at the .05 level. None of the other main variables and interactions were significant for Learning Total. A significant (.05 level) sex difference also appeared for Learning Rate, with the women *Ss* showing a higher average rate. The interaction of Extraversion and Neuroticism was quite insignificant ($F = .38$). None of the other main effects or interactions approached significance for Learning Rate. Inspection of mean Learning Rate for the MPI subgroups given in Table 4 shows little consistency in the patterns of the means for the two sex groups, and the interactive effect of Extraversion and Neuroticism shown in Table 2 is not evident in Table 4.

DISCUSSION

The results of this replicated experiment provide little evidence that the personality dimensions of Extraversion and Neuroticism, as defined by the Maudsley Personality Inventory, are related to the motor learning performance measured by the inverted alphabet printing task. The interactive effect of Extraversion and Neuroticism upon Learning Rate did not appear when the original study was replicated, and these contradictory results can probably be attributed to the relatively low reliability of the Learning Rate measure. The psychometric inadequacy of Learning Rate as a measure of performance is also evident in the significant sex difference found for this variable in Experiment II when compared to the absence of any evidence of a significant sex difference in Experiment I. Learning Total appears to be a more consistent measure of performance; the reliability coefficients were quite high, and the significant sex difference appeared in both samples. Since the MPI variables were not significantly related to Learning Total in either experiment, the conclusion that the personality and learning variables are not related seems justified. It is possible that the 10-min. practice period was not long enough to generate a sufficient amount of reactive inhibition to demonstrate a significant difference in learning between the groups of *Ss* dichotomized on the Extraversion scale. Arguing against this possible inadequacy is the evidence from previous studies that a 10-min. performance period results in sufficient reactive inhibition as shown by reminiscence indices. Another possibility is that simple dichotomization of *Ss* on the MPI scales does not provide groups that are extreme enough to demonstrate significant relationships between temperament traits and learning. Selecting *Ss* at the extremes of the MPI continua, e.g., upper and lower quarters, might result in significant differences in learning.

SUMMARY

The Extraversion and Neuroticism scales of the Maudsley Personality Inventory were administered to 72 college students who were dichotomized on the basis of their scale scores. The inverted alphabet printing task was used to provide measures of motor learning performance. Analyses of variance of the motor learning scores showed that the women learned significantly faster than the men, but the main effects of Extraversion and Neuroticism

were not significant. The interaction of the two personality variables approached statistical significance (.10 level). Consequently, the study was replicated with another group of 99 similar *S*s. The sex difference again was significant for the replication group, and the Extraversion and Neuroticism main effects were again not significant. The interactive effect of Extraversion and Neuroticism with the learning measures was far from significant in the replication. It was concluded that there was no evidence that the MPI personality variables were related to performance on this motor learning task.

REFERENCES

ALBRIGHT, L. E., BORRESEN, C. R., & MARX, M. H. Reactive inhibition as a function of same-hand and opposite-hand intertrial activity. *J. exp. Psychol.*, 1956, **51**, 353–357.

ARCHER, E. J. Postrest performance in motor learning as a function of prerest degree of distribution of practice. *J. exp. Psychol.*, 1954, **47**, 47–51.

ARCHER, E. J., & BOURNE, L. E., JR. Inverted-alphabet printing as a function of intertrial rest and sex. *J. exp. Psychol.*, 1956, **52**, 322–328.

BENDIG, A. W. Extraversion, neuroticism, and manifest anxiety. *J. consult. Psychol.*, 1957, **21**, 398.

EYSENCK, H. H. Reminiscence, drive, and personality theory. *J. abnorm. soc. Psychol.*, 1956, **53**, 328–333.

EYSENCK, H. J. *The dynamics of anxiety and hysteria.* New York: Praeger, 1957.

JENSEN, A. R. Personality. *Annu. Rev. Psychol.*, 1958, **9**, 295–322.

KIENTZLE, M. J. Properties of learning curves under varied distribution of practice. *J. exp. Psychol.*, 1946, **36**, 187–211.

KIMBLE, G. A. An experimental test of a two-factor theory of inhibition. *J. exp. Psychol.*, 1949, **39**, 15–23. (a)

KIMBLE, G. A. Performance and reminiscence in motor learning as a function of the degree of distribution of practice. *J. exp. Psychol.*, 1949, **39**, 500-510, (b)

MOSIER, C. I. Batteries and profiles. In E. F. Lindquist (Ed.), *Educational measurement.* Washington: American Council on Education, 1951. Pp. 764–810.

RUCH, F. L., & WARREN, N. *Working with psychology.* (1st ed.) Chicago: Scott, Foresman, 1941.

SCHUCKER, R. E., STEVENS, L. B., & ELLIS, D. S. A retest for conditioned inhibition in the alphabet printing task. *J. exp. Psychol.*, 1953, **46**, 97–102.

STANLEY, J. C. Fixed, random, and mixed models in the analysis of variance as special cases of finite model III. *Psychol. Rep.*, 1956, **2**, 369.

THORNDIKE, R. L. *Personnel selection.* New York: Wiley, 1949.

WALKER, HELEN M., & LEV, J. *Statistical inference.* New York: Holt, 1953.

WASSERMAN, H. N. The effect of motivation and amount of prepractice upon inhibitory potential in motor learning. *J. exp. Psychol.*, 1951, **42**, 162–172.

(Received September 22, 1958)

SOME RELATIONSHIPS BETWEEN MANIFEST ANXIETY, AUTHORITARIAN TENDENCIES, AND MODES OF REACTION TO FRUSTRATION

ROBERT D. SINGER[1] AND SEYMOUR FESHBACH

University of Pennsylvania

THE intent of the present study is to obtain data bearing upon the relationship between psychopathology and authoritarianism in a group of middle-class college students by analyzing the correlations between authoritarianism, manifest anxiety, and self perceived reactions to frustration.

Studies dealing with the relationship between authoritarian tendencies and psychopathology have yielded conflicting results (Gough, 1951; Jones, 1953; Davids, 1955; Freedman, Webster & Sanford, 1956) and the issue has elicited considerable controversy (Masling, 1954; Frenkel-Brunswik, 1954). It is rather fruitless to argue what the correlation *should* be since it is apparently possible to draw diverse inferences regarding the nature and extent of the relationship, depending upon which statements made by the authors of *The Authoritarian Personality* (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950) one chooses to stress.

Masling (1954, p. 316) contends that there is a bias in favor of seeing the equalitarian personality as being "a fairly well adjusted individual" and the authoritarian as "a nasty fellow." However, Frenkel-Brunswik (1954) points out that while high scorers may suffer from strong repressive tendencies, the low scorer may engage in excessive self-criticism and introspection as well as manifest other neurotic features.

The question, then, of the connection between authoritarian tendencies and variables usually associated with various states of mental health becomes an empirical issue that can only be resolved by the accumulation of pertinent data. Much of the existing evidence has been conflicting, however, partly as a result of methodological insufficiencies such as neglect of response set artifacts, and, of more importance, of the assumption that the relationship between mental health and authoritarianism should be invariant. There is sufficient information available regarding the multidimensionality of the F scale, the principal measure of authoritarian tendencies in current use, to lead one to anticipate markedly different relationships for different social groups. Thus a high F score may in part be the result of rigid, repressive personality defenses and in part of membership in a group whose values and standards are authoritarian. The prevailing ideology of the social group being sampled and the extent to which group members conform to its standards should have a decided effect upon the correlation of mental health with authoritarian tendencies: the greater the impact of group ideology, the lower the correlation. Other factors such as the kind of measure of mental health used and the particular environmental circumstances prevailing when the measures are obtained should also affect the correlation; e.g., if high authoritarians can readily adjust to the authoritarian structure presented by military service, then it is not surprising to find an insignificant correlation between authoritarian tendencies and a measure of personal adjustment in a sample of naval personnel (Masling, 1954).

A third factor that may affect the correlation obtained between authoritarian tendencies and measures of mental health is the type of measure of adjustment used. Self-report scales, in particular, may be strongly influenced by the way in which an authoritarian individual perceives himself in relation to other people. The present study focuses on this problem.

METHOD

One hundred forty-seven male undergraduates, drawn from laboratory sections of the introductory psychology course at a large Eastern metropolitan university, were administered a questionnaire composed of the Taylor Manifest Anxiety Scale (Taylor, 1953), the California F Scale of authoritarianism (Adorno,

[1] Now at Eastern Pennsylvania Psychiatric Institute, Philadelphia, Pa.

et al., 1950),[2] and a measure of various modes of reaction to frustration. The latter scale, developed by Child and Waterhouse (1953), yields measures of intrapunitiveness, extrapunitiveness, and constructiveness in response to frustration. The questions are so constructed that the *S* rates himself in comparison to how he thinks the average person behaves. Each of the subscale scores is based on a different set of 15 items so that it is possible for an *S* to obtain high scores on all three measures.

The scoring procedures employed for each of the tests were as follows: (*a*) For the F scale, a score ranging from 7 to 1 (4 being omitted) was assigned to each of the six alternatives such that the higher the score, the more authoritarian the response. For any particular *S*, his F scale score consisted of the mean item score for the entire set of items. (*b*) An *S*'s score on the Taylor scale consisted of the number of items answered in the direction indicating anxiety. (*c*) There are six alternatives to each item on the Child-Waterhouse scales. A score ranging from 1 to 6 was assigned to each item such that a higher score indicated a higher degree of the response tendency being measured. For any item, a score of 3 or under indicated that the respondent felt that the behavior described in this item was less true for him than for the average person, while a score of 4 or over indicated that this behavior was more characteristic of him than of the average person.

RESULTS

A Pearson correlation was computed between the F scale and the Taylor scale. The correlation, as shown in Table I, proved to be +.34, significant at the .01 level. To check for possible influence of acquiescence set, the items of the Taylor scale that yield an anxiety score when answered "false" were separately scored. In the version of the Taylor scale used there were 11 such items. A Pearson *r* was computed and found to be .15 which is not significant for an *N* of 142. However, the 11 items form a severely restricted range in comparison to the original 50-item scale. When corrected for restriction of range, the correlation rose to a significant figure of .33 which is almost identical with the original correlation. As a further check, a chi square was also computed by dichotomizing at their respective medians both the F scale distribution and the shortened Taylor scale distribution based on the 11 items yielding an anxiety score when marked false. A chi square of 3.9 significant at the .05 level was obtained. While this finding does not demonstrate that acquiescence set had no influence upon the correlation between the F scale and the complete Taylor scale, it does show that there is a significant positive relationship between authoritarianism and manifest anxiety when the possible effect of acquiescence set is eliminated.

[2] Form 45-40 of the F scale was used. This form includes the following individual items: 1, 2, 4, 6, 8, 9, 12, 13, 16, 18, 19, 21, 22, 23, 25, 26, 27, 29, 31, 33, 34, 35, 37, 38, 39, 41, 42, 43, and 44.

TABLE 1

RELATIONSHIP BETWEEN AUTHORITARIAN TENDENCIES, MANIFEST ANXIETY, AND REACTION TO FRUSTRATION

	N	*r*	*P*
Authoritarian scale/Taylor scale	147	+.34	.01
Authoritarian scale/Intrapunitive scale	147	+.50	<.01
Authoritarian scale/Extrapunitive scale	147	+.23	<.01
Authoritarian scale/Constructiveness scale	147	−.20	<.05
Taylor scale/Intrapunitive scale	147	+.51	<.01
Taylor scale/Extrapunitive scale	147	+.17	<.05
Taylor scale/Constructiveness scale	147	−.49	<.01

A further analysis of the F scale and the complete Taylor scale was undertaken by dividing the 147 cases into three subgroups, the high authoritarians consisting of the upper quartile, the middle authoritarians consisting of the second and third quartiles, and the low authoritarians consisting of the fourth quartile. The results of this analysis show the anxiety scores of low and middle authoritarians to be almost identical. Thus, for a considerable range of the F scale, there is no association between F and Taylor scale scores. The obtained correlation results primarily from the correspondence between high authoritarianism and high anxiety.

The correlations of the F and Taylor scales with the measures of reaction to frustration are also presented in Table 1. While the correlations vary in size they are all statistically significant. The F scale correlates positively with the intrapunitive and extrapunitive measures and negatively with the measure of constructive reactions to frustration. Similar results were obtained for the Taylor scale, the only appreciable difference being the greater magnitude of the negative correlation of the Taylor scale with the constructiveness measure. It may be of interest to note that these results were obtained despite the quite small variance of *S*s on the Child-Waterhouse scales.

Table 2 shows the same pattern of results

TABLE 2
REACTIONS TO FRUSTRATION OF HIGH AND LOW ANXIOUS, AND HIGH AND LOW AUTHORITARIAN SUBJECTS

Range:	Highs (20–38)		Middles (9–19)	Lows (1–8)		t	p	Scores of Low Anxious *Ss* are:
	M	σm	M	M	σm			
Taylor scale (M—14.1; SD—6.8)								
Reactions to frustration								
Extrapunitive	52	1.4	50	48	1.2	2.2	<.05	less extrapunitive
Intrapunitive	55	1.3	51	48	.9	4.3	<.01	less intrapunitive
Constructive	51	1.2	54	57	1.1	3.5	<.01	more constructive
Range:	Highs (4.7–6.0)		Middles 3.0–4.6)	Lows (1.2–2.9)		t	p	Scores of Low Authoritarian *Ss* are
	M	σm	M	M	σm			
Authoritarian scale (M—4.0; SD—1.8)								
Reactions to frustration								
Extrapunitive	54	1.4	50	47	1.3	3.7	<.01	less extrapunitive
Intrapunitive	55	1.4	52	49	1.4	3.1	<.01	less intrapunitive
Constructive	51	1.6	53	56	1.2	2.5	<.05	more constructive

when we examine the mean differences on the frustration scales between *Ss* high, middle, and low in authoritarianism and those high, middle, and low in anxiety. In contrast to the relationship between the middle ranges of F scale scores and Taylor scale scores, the scores of the middle quartile groups on reactions to frustration consistently fall between those of the upper and lower quartiles. Thus it appears that the relationship between authoritarian tendencies and reactions to frustration is more uniform than the corresponding relationship between authoritarian tendencies and manifest anxiety.

Because the scale measuring reactions to frustration is so constructed that the *S* rates himself in comparison to how he thinks the average person reacts, it is possible to put the the following question to the data: How do *Ss* high and low in authoritarianism and in anxiety perceive themselves in relation to the average person? The results of this analysis are presented in Table 3. An *S* who perceived his reactions as being characteristic of the average person would obtain a mean item score of 3.5 or, over the 15 items dealing with a particular mode of reaction to frustration, a hypothetical average score of 52.5. On none of the measures of the reactions to frustration are the means of the *Ss* high in anxiety or authoritarianism significantly different from the hypothetical mean that would be obtained if the *S* perceived himself as like the average person in regard to these traits. However, in all cases, the means of the *Ss* low in authoritarianism and anxiety are significantly different from the hypothetical average. They see themselves as being less intrapunitive, less extrapunitive, and more constructive in their reactions to frustration than the average person.[3]

The direction of the correlations obtained between authoritarian tendency and mode of reaction to frustration warrants comment. The scales of extrapunitiveness and intrapunitiveness are both measures of maladaptive reactions, one consisting of irrational hostility

[3] Before concluding that *Ss* high in anxiety and authoritarianism see themselves as more like the average person than do those who score low on these variables, it was necessary to consider the possibility that the means obtained by the Highs may have been the resultant of averaging a highly variable set of scores—such that a large number fall considerably below the mean and a large number considerably above the mean. The distribution of total scores and the distribution of item scores for each *S* were examined. Both the interindividual and the intra-individual analysis indicated that Highs on each variable rate themselves as more similar to the average person than do Lows.

TABLE 3
DEVIATIONS OF HIGH AND LOW ANXIOUS AND HIGH AND LOW AUTHORITARIAN SUBJECTS FROM A HYPOTHETICAL AVERAGE PERSON SCORE

Deviations to Frustration	Hypothetical Average Score (Av)	High Anxiety				Low Anxiety			
		M	σm	*d*(*M*-*av*)	*p*	*M*	σm	*d*(*M*-*Av*)	*p*
Extrapunitive	52.5	52	1.4	−.5	>.05	48	1.2	−4.5	<.01
Intrapunitive	52.5	55	1.3	+2.5	>.05	48	.9	−4.5	<.01
Constructive	52.5	51	1.2	−1.5	>.05	57	1.1	+4.5	<.01
Reaction to frustration	**Hypothetical Average Score (Av)**	**High F**				**Low F**			
		M	σm	*d*(-*MAv*)	*p*	*M*	σm	*d*(*M*-*Av*)	*p*
Extrapunitive	52.5	54	1.4	+1.5	>.05	47	1.3	−5.5	<.01
Intrapunitive	52.5	55	1.4	+2.5	>.05	49	1.4	−3.5	<.05
Constructive	52.5	51	1.6	−1.5	>.05	56	1.2	+3.5	<.01

towards other people and objects, the other of excessive self-blame and extreme lack of self-confidence. The third scale measures the degree of constructiveness and adequacy in the face of frustrating situations. In considering the possible reactions of authoritarians to frustration, some comments of Sanford (1950, p. 47) are pertinent. He states that equalitarians "seem inclined to take a rational and impunitive position in a situation of social stress, and they demonstrate a tendency to be relaxed in the face of personal inconvenience. . . ." The present data indicate that the lower a person's score on authoritarianism the less likely he is to rate himself as aggressive and to report feelings of self-blame and inadequacy. And while the high scorer obtains both higher intrapunitive and extrapunitive scores than the low scorer, intrapunitive reactions are more predominant. These data lend little support to the frequent characterization of the authoritarian as an extremely aggressive, somewhat hysterical personality in contrast to the more self-doubting, insecure but tolerant equalitarian (Frenkel-Brunswik, 1954). On each of the scales used in this study, authoritarian tendencies are positively associated with neurotic or psychopathological tendencies.

The analysis of the questionnaire on reactions to frustration in terms of comparisons with the average person provides additional insight into the personality dynamics of high and low authoritarians: The high authoritarians rate themselves high on pathological states and behaviors while they view their reactions to frustration as being quite average. Apparently, high authoritarians can recognize their own behavior when asked to rate concrete instances of such behavior. On the other hand, high authoritarians tend to assume that their behaviors are just like those of other people when, in fact, as in the case of the Taylor scale, they are not. Thus, despite the fact that authoritarians may score higher than the average on self-perceived measures of neuroticism, they nevertheless may fail to perceive their deviation from the norm. These results are consistent with a series of findings that high authoritarians tend to perceive their peers as having attitudes and personality characteristics similar to their own (Scodel & Mussen, 1953; Scodel & Freedman, 1956; Crockett & Meidinger, 1956; Rabinowitz, 1956).

The question of how high and low authoritarians actually do react to frustrating situations still remains. Child and Waterhouse (1953) have presented some evidence for the validity of their scale; however, the veridicality of the perceptions of high and low authoritarians needs to be more firmly established. The findings indicate the necessity for a more careful delineation of the authoritarian syndrome, particularly in regard to the aetiological connections of anxiety and authoritarianism. The findings also point up the value of examining the correspondence between the self-perceptions of high and low authoritarians and their objectively observed reactions.

SUMMARY

A sample of 147 college students were administered a questionnaire composed of the Taylor Manifest Anxiety Scale, the California F Scale and the Child-Waterhouse reactions to frustration scale. This latter scale provides indices of intrapunitive, extrapunitive, and constructive responses to frustration.

A significant but low positive correlation was obtained between the F scale and the Taylor scale, the correlation being primarily accounted for by the association between high scores on the F scale and high scores on the Taylor scale. An analysis of the data in which the possible effects of "acquiescence set" were eliminated still yielded a significant, positive relationship. In addition both the Taylor and F scales correlated positively with measures of both intrapunitive and extrapunitive tendencies and correlated negatively with the index of constructive reactions to frustration. The data further indicate that although highly anxious and highly authoritarian *S*s deviate from the norm on the Taylor scale, they perceive themselves as being similar to the average person in the manner in which they react to frustration. On the other hand, low anxious and low authoritarian *S*s perceive themselves as being more constructive, less intrapunitive, and less extrapunitive than the average person.

REFERENCES

ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The Authoritarian Personality*. New York: Harper, 1950.

CHILD, I. L., & WATERHOUSE, I. K. Frustration and the quality of performance: III. An experimental study. *J. Pers.*, 1953, **21**, 298–311.

CROCKETT, W. H., & MEIDINGER, T. Authoritarianism and interpersonal perception. *J. abnorm. soc. Psychol.*, 1956, **53**, 378–380.

DAVIDS, A. Some personality and intellectual correlates of intolerance of ambiguity. *J. abnorm. soc. Psychol.*, 1955, **51**, 415–419.

FREEDMAN, M., WEBSTER, H., & SANFORD, N. A study of authoritarianism and psychopathology. *J. Psychol.*, 1956, **41**, 315–322.

FRENKEL-BRUNSWIK, ELSE. Social research and the problem of values: A reply. *J. abnorm. soc. Psychol.*, 1954, **49**, 466–471.

GOUGH, H. Studies of social intolerance: II. A personality scale for anti-Semitism. *J. soc. Psychol.*, 1951, **33**, 247–255.

JONES, M. B. Aspects of the autonomous personality: I. Manifest anxiety. *USN Sch. Aviat. Med. Res. Rep.*, 1953, Proj. No. NM 001 058.25.03.

MASLING, M. How neurotic is the authoritarian? *J. abnorm. soc. Psychol.*, 1954, **49**, 316–318.

RABINOWITZ, W. A note on the social perceptions of authoritarians and nonauthoritarians. *J. abnorm. soc. Psychol.*, 1956, **53**, 384–386.

SANFORD, F. H. Authoritarianism and leadership: A study of the follower's orientation to authority. Philadelphia: Inst. Res. Human Relations, 1950.

SCODEL, A. & FREEDMAN, M. L. Additional observations on the social perceptions of authoritarians and nonauthoritarians. *J. abnorm. soc. Psychol.*, 1956, **52**, 92–95.

SCODEL, A. & MUSSEN, P. Social perceptions of authoritarians and nonauthoritarians. *J. abnorm. soc. Psychol.*, 1953, **48**, 181–184.

TAYLOR, JANET A. A personality test for manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

TITUS, H. E. & HOLLANDER, E. P. The California F Scale in psychological research. *Psychol. Bull.*, 1957, **54**, 47–64.

(Received September 23, 1958)

DISCRIMINATION AND REVERSAL LEARNING IN IMBECILES

N. O'CONNOR AND B. HERMELIN[1]

Medical Research Council, Institute of Psychiatry, London, England

THEORETICAL DISCUSSION

THE learning difficulties of imbeciles affect a wide range of activities. It seems justified, nevertheless, to regard processes which require the utilization of general classifying principles and the manipulation of symbols—such as words, letters, and numbers—as more severely impaired than motor performances. The reason for this may conceivably be found in the role played by language in the acquisition of different abilities. Whereas motor skills may be acquired by a process akin to conditioning, problem solving may be to some extent dependent on formulating a principle of solution. In the former, a verbal formulation is unnecessary. In the latter, it may be essential. In addition to sheer weakness in the development of speech functions, it has been proposed by Luria (1956) and other Soviet psychologists that the severely subnormal suffer from a dissociation between verbal and motor systems.

Luria's view is that while at the beginning of normal development speech serves mainly to communicate with other people, it gradually begins to assume a directive function; once the age of about 5½ is reached, internal speech and verbal self-instruction guide and anticipate motor behavior. New connections formed with the participation of the verbal system can be abstracted and generalized, reflecting a fuller picture of the environment as words become a means of systematizing experience. From this age onwards, Luria claims that verbal connections tend to antedate those built up by motor acts. In this sense, the verbal system regulates direct actions, and this regulation forms the basis of conscious, voluntary behavior. Connections between stimuli, if they are formed through verbal associations, enable the child to abstract and generalize the significant elements in the signals and to orient himself among them.

[1] The authors wish to thank J. M. Crawford, Physician Superintendent of Botleys Park Hospital, and his staff, who kindly made available the facilities for this experiment.

In the severely subnormal child, language does not develop sufficiently to assume this orientating function. For such children, words do not play a regulating role but are at best impelling signals, and the result is a lack of connection between verbal and motor behavior. Defective children of five or six years have difficulty in carrying out simple verbal instructions, although they are able to repeat such instructions correctly. Luria thinks that such response connections which are developed without verbal associations are extremely unstable, depend on constant reinforcement, and disintegrate at a slight change in the manner of presenting the signals. He states that as the system of verbal abstractions does not take its necessary role in the formation of new links, the subnormal child's psychological development must inevitably and increasingly suffer. He adds that the rational way to strengthen the joint operation of the two signal systems is to change to an open, straightforward bringing together of verbal and visual signals. At the same time, Luria holds that in severe cases of defect, an inhibiting state of the whole cortex exists, which may render such treatment ineffective. This kind of theory accords a very important role to words and symbols in the development of mental defectives and makes verbal deficit primary.

The experiment reported here was designed to test an aspect of this theory, using a discrimination and reversal task, and to elucidate further the relation between words and sensory-motor behavior in imbeciles.

PREVIOUS REVERSAL EXPERIMENTS

In reversal experiments a stimulus choice is usually rewarded until it is consistently made, whereupon the reward is shifted to the previously negative stimulus. Transfer in such a situation should be negative, and the subject (*S*) begins to learn the reversal task with a considerable handicap. In the same manner as a positive stimulus acquires excitatory response tendencies during training, a negative one gives rise to inhibition. This inhibition would

first have to be dispersed before new excitatory tendencies towards the formerly negative stimulus could become effective.

In reversal experiments with the feeble-minded, no significant differences between them and normals of comparable mental age have usually been found in the rate of discrimination and reversal learning. Gardner (1948), who trained his *S*s to choose one out of three differently marked boxes, found a decrement in the number of correct responses for all IQ levels when the reward was shifted to a previously incorrect box. Idiots seemed to have had greater difficulties in this task than the other subnormal *S*s. Plenderleith (1956) required her *S*s to learn to choose one of two stimuli in a consistent fashion; after this discrimination was learned, the *S*s were required to shift their responses to the previously incorrect stimulus. The feeble-minded children did not differ significantly from the normals in either discrimination trials or in the reversal scores for trials given after 24 hours. Stevenson and Zigler's (1957) *S*s were trained to select one out of three stimuli; following different degrees of training, they were switched to a new size discrimination involving the same stimuli. Normal children, feebleminded children, and feebleminded adults did not differ significantly in the acquisition of the first solution, nor in their performance on the second problem. No significant differences appeared as a function of degree of training. Kounin (1948) reported that feeble-minded *S*s reversed a motor response faster than normals. His *S*s were, however, *instructed* to reverse the response, and it is questionable if his experimental conditions made a comparison with other reversal experiments meaningful. He interpreted his results in terms of such concepts as "functional rigidity" and "a high barrier between psychological regions," which prevent interference from previously acquired response patterns. The same concept of rigidity could, of course, be applied to explain exactly opposite behavior manifestations. Less mobile systems might result in difficulty for the *S* to reverse an established discrimination response. No controlled reversal experiments with imbecile children seem to be reported in the literature.

TABLE 1
MEAN ACQUISITION AND REVERSAL SCORES FOR NORMALS AND IMBECILES

	Discrimination	Reversal
First Experiment		
Normals	31.5 ± 24.9	25.4 ± 16.6
Imbeciles	39.6 ± 24.7	11.0 ± 5.2
Second Experiment		
Imbeciles	25.1 ± 13.7	39.6 ± 27.0

FIRST EXPERIMENT

Method

Black squares of different sizes were used as stimuli. They were presented, two at a time, on a tray on which they could be moved to reveal a reward. The squares were painted on red cardboard, each card measuring 20 x 20 cm. The size of the squares differed by a ratio of 5:8. Candy was given as reward for every correct response. After the trials had been completed, each child was asked to give a verbal account of the principle according to which he made his choices.

In the discrimination trials, half the *S*s in each group were rewarded for choosing the smaller stimulus; the others were rewarded for choosing the larger of the two squares. Trials continued until 10 consecutive correct responses occurred. Immediately after the criterion for discrimination was reached, and without any comment by the experimenter, the *S*s were given a number of reversal trials in which the formerly negative of the two stimuli was the correct rewarded choice. The reversal trials continued until 10 correct consecutive choices were made.

Ten institutionalized imbecile children and 10 normal children acted as *S*s. The imbeciles had a mean IQ of 42 (range, 33–50), a mean CA of 11.5 years (range, 9–15), and a mean MA of 4.9 years (range, 3.4–6.7). The control group consisted of normal school children with a mean CA of 5.1.

Results

The imbeciles, on the whole, needed more trials before reaching the required criteria. There is, however, as indicated in Table 1, a considerable overlap between the individual scores in the two groups, and intragroup

variances are large. Imbeciles needed from 10–86 responses, mean 39.6; the normals required 0–73 responses, mean 31.5, before they could discriminate consistently. The difference between those scores as measured by *t* was, however, nonsignificant.

The crucial difference between the two groups in this experiment is evident in the reversal trials, in which the imbecile group performed better than the normals. While the imbeciles needed only a mean of 11 trials to reverse successfully a previously learned response, the normals required a mean of 25.4 trials. This difference is even more pronounced if one relates it to the original discrimination scores of the two groups. While the normals need nearly as many trials again for reversal as for discrimination, the imbeciles need less than a third of the number of trials for reversal compared with discrimination. The difference between normal and imbecile reversal scores is significant at the .01 level, and while no significant difference is apparent between discrimination and reversal scores in the normal group, this difference reaches a significance level of .05 for the imbeciles.

Eight of the 10 normal children could formulate the relevant verbal hypothesis of size discrimination either spontaneously or when asked, whereas in no instance was the experimenter able to elicit the principle from any of the imbecile children, excepting one child who stated it spontaneously.

Discussion

As a tentative explanation of this finding, it was hypothesised that the normal child tends to formulate the solution to the problem verbally: for example, "the candy is always under the big square." In the reversal task, this verbal self-instruction comes into conflict with the new reinforcement and interferes with the solution of the task. What seems to be required of the child is some unlearning of a previously established response tendency before new learning can take place.

The imbecile child, on the other hand, supposedly never formulated a verbal hypothesis; therefore, he might be able to respond more immediately to the reinforcement coming from the direct stimuli. According to Spence (1940), even an unlearned or only partly learned discrimination should lead to an interference with subsequent reversal. Because one stimulus had been reinforced, it should have led to the building up of excitatory response tendencies, whereas the nonreinforced response should have been inhibited. Response tendencies formed without the participation of the verbal system might, therefore, be less stable and extinguish faster than those formed with the additional verbal formula. If the inability of imbeciles to use words as guiding hypothesis was a determining factor in the results, training and reinforcing such verbalization should slow down the reversal process and bring the scores nearer to those obtained by the normal children.

SECOND EXPERIMENT

Method

The discrimination-reversal experiment was repeated with another group of 10 imbeciles, matched in age and IQ with the previous one. Whereas no comments were made with the previous groups during the trials, however, each correct discrimination response was verbally reinforced with the second group. Whenever the child responded correctly, he was asked, "Under which picture did you find the candy? Under the big or under the little one?" If the response was "this," with pointing, the next question was, "Which one is this?" On the response "the big" or "the little one," it was again asked, "Tell me again now, under which picture was the candy?" Questioning persisted until the child gave the correct verbal account, and the same procedure was repeated after every correct motor response.

No questions were asked or comments made during the reversal trials, which, as previously, followed immediately after the criterion in the discrimination task was reached.

Results

In the acquisition task, this second group needed a mean of 25.1 to reach the required discrimination criterion, compared with 39.6 trials for the imbeciles and 31.5 for the normals in the previous experiment. These differences are, however, statistically nonsignificant.

In the previous experiment, normal reversal

and discrimination scores were statistically equal, and the imbeciles needed significantly *fewer* trials for reversal than for discrimination. The second imbecile group, on the other hand, needed a larger number of trials to reverse than it needed to acquire the preceding discrimination response. As was the case with the normal children, the difference between discrimination and reversal scores was not significant for the second group of imbeciles. Their reversal scores differ significantly from those of the first group ($p = .01$) and are not significantly different from those of the normals.

At the beginning of the test, little connection between verbal and motor behavior of the *S*s was evident. Thus, for instance, after responding correctly, a child would state, equally correctly, that he had found the candy under the larger square. He would then choose the smaller one at the next trial, and finding no reward, would say, "It is under the big one." Yet the next response would again be incorrect, although the child was quite able to point to the larger stimulus square when asked to do so. However, the connections between the two kinds of response were gradually formed, and the children tended to advance a hypothesis as to where the reward would be before making the motor response. They would spontaneously say, "It is under the small one," and would then make the correct choice.

When the stimulus situation was reversed, the previously appropriate verbal response tended to persist as a stereotype, although it was no longer an adequate guide to motor behavior. Nevertheless, the *S*s, although still using previously learned and now inappropriate verbal regulations, gradually reversed their choice behavior, responding to the direct signals. But while reversing to the newly correct smaller stimulus, they still would say, "It is under the big one," although again they were quite able to point to the smaller square if requested to do so. Only three children were able to reverse their verbal response—two simultaneously with and one gradually following the motor ones. Four gradually stopped any verbal comment, and three persisted in the verbal stereotype previously acquired until the end of the trials. For those *S*s, the verbal system seemed to have lost its regulating, orienting role in relation to the direct signals.

Discussion

Pavlov assumed the existence of two signal systems in the human organism, one governing motor and the other verbal behavior. In the normal child, the second system develops to dominate and guide response to almost all direct signals. If one assumes with Luria and his colleagues that in certain forms of disease either lack of development in one of the systems or lack of fusion between the two is a prominent symptom, then decreased mobility as well as instability of certain responses become explicable. Applied to the present investigation this formulation suggests that while conditioning with the influence of the verbal system takes place more quickly than without it, the connections thus developed are less mobile than those established on the basis of direct signals alone.

The experiment seems to show that the imbeciles, as long as they are not verbally reinforced, learn a perceptual motor habit, whereas the normals acquire two habits, one perceptual motor and one verbal. It must be borne in mind that the mean age of the normal group in this experiment was just five years, the stage of development at which, according to Luria, verbal self-regulation of behavior is just beginning. Older children would presumably be able to adapt their verbal as well as their motor behavior more to a changed stimulus situation with more facility. However, with this particular age group of normals, as well as with imbecile children, words seemed to be a powerful determinant preventing them from reversing "set" rapidly. Conversely, lack of verbal regulation of motor activity in imbeciles seemed to result in unstable responses which could be rapidly reversed. A verbal habit, once set up in accordance with the motor one, lacks mobility and slows down reversal.

SUMMARY

Imbecile children matched for mental age with normal five-year-olds learned a size discrimination problem with more difficulty than the normals, but showed greater facility in the appropriate reversal task. An explanation of this phenomenon in terms of the verbal regulation of motor behavior was offered and tested.

REFERENCES

GARDNER, P. L. The learning of low grade aments. *Amer. J. ment. Def.*, 1948, **50**, 59–80.

KOUNIN, J. S. The meaning of rigidity: A reply to Heinz Werner. *Psychol. Rev.*, 1948, **55**, 157–160.

LURIA, A. R. *Problems of the higher nervous activity of normal and abnormal children*, Moscow: Acad. Pedagog. Sci., R.S.F.S.R., 1956.

PLENDERLEITH, M. Discrimination learning and discrimination reversal learning in normal and feebleminded children. *J. genet. Psychol.*, 1956, **88**, 107–112.

SPENCE, K. W. Continuous versus non-continuous interpretation of discrimination learning. *Psychol. Rev.*, 1940, **47**, 271–288.

STEVENSON, H. W., & ZIGLER, E. F. Discrimination learning and rigidity in normal and feebleminded individuals. Unpublished report, Univer. of Texas, 1957.

(Received September 23, 1958)

IDENTITY DIFFUSION IN LATE ADOLESCENTS[1]

GORDON W. BRONSON[2]
University of California

In a recent paper on identity, Erikson (1956) has described late adolescence as a period of "identity crisis," defining it as a developmental stage in which individuals are involved in a process of personal redefinition—sometimes with an intense self awareness, sometimes on the level of less conscious processes. Throughout the adolescent period, a combination of intrapsychic and cultural forces act to initiate and aid this process. Experiences of heterosexual intimacy reactivate those latent aspects of the self which are products of intimacy experiences of early childhood. A heightened sense of personal independence leads to new relationships in an enlarged social environment, giving new perspectives on past relationships. Social pressures to make lifelong commitments in occupation and marriage call for reappraisal of social goals and interpersonal ideals. The effect of these forces is to present the individual with a complex and at times confusing array of self-images—some coming from the reactivated past, some from a broader perspective of the immediate future.

Erikson presents several clinical cases illustrating the bizarre and confused behavior which can at times accompany the identity crisis. He uses the concept of "identity diffusion" to refer to the psychological chaos experienced by these adolescents. While his illustrations come from psychopathology, in a less striking form identity diffusion can be expected to be a frequent characteristic of the normal adolescent passing through the period of identity crisis. In the process of reappraising his past, evaluating personal assests and liabilities, reformulating personal ideals and goals and overtly testing new interpersonal roles, the adolescent will at times be uncertain as to which of the diverse possibilities is most truly he. Lacking a stable sense of self, he experiences fluctuating evaluations of his social environment and wide variations in his interpersonal behavior. Identity diffusion thus represents a condition of psychological fluidity and of weakened repressions. It is not surprising that this condition should result in an increase in inner tensions and a heightened sense of anxiety.

The present study aims at demonstrating the existence of interrelationships among various manifestations of identity diffusion. Four salient characteristics derived from Erikson's clinical descriptions were chosen for measurement. In contrast with persons with a stable sense of self, those in a state of identity diffusion should: (*a*) be less certain about the relationship between the past and current notions of self, (*b*) show a higher degree of internal tension or anxiety, (*c*) be less certain about dominant personal characteristics, and (*d*) fluctuate more in their feelings about self. A finding of significant correlations among measures relating to different aspects of identity diffusion will give empirical support to this construct.

Method

Sample

Measures relating to the four qualities selected for study were obtained on a sample of 46 volunteer college students. All but seven were within the 19 to 22 age range and all but four were women. Since this age group represents the end of the adolescent period it can be expected to show appreciable variation in the degree of identity diffusion rendering it thus optimal for the correlational approach used in this study.

Techniques of Measurement

Interview measures. Two of the four measures relating to identity diffusion were based on interview material. The interviews, lasting about 20 min., covered areas of self evaluation, aspirations and ideals, reactions to other people, and childhood peer and parent relationships. The semistructured interviews were conducted by a clinically trained psychologist who was otherwise unknown to the *S*s. Under these circumstances, and with anonymity assured, *S*s cooperated comfortably in the interview situation. Since the purpose was to assess the degree to which formulations in

[1] This report is based on a dissertation submitted to the Graduate School of the University of California, Berkeley, in partial fulfillment of the requirement for the degree of Doctor of Philosophy.

[2] Now at Mills College.

these areas represented an integrated basis for a conception of self, little attempt was made to probe into significant omissions or apparent contradictions. Of the total of 46 *Ss*, 18 randomly selected individuals were interviewed. Correlations involving the two measures based on interviews are thus based on an *N* of 18.

The interviews were tape recorded and later submitted to three clinical psychologists for rating.[3] Judgments were made on the two following six point scales:

Continuity with the Past. Raters based their judgments on the following instructions: "Does the person's conscious evaluation of himself seem to have evolved as an integrated expression of past experiences? Is there a sense of depth and permanence to the self-descriptions or might the individual as effectively take on different self-notions should circumstances so demand?"

This rating reflects the judges' general impression of the stability and rootedness of the interpersonal values and techniques expressed by the *Ss* in their interview behavior. High ratings represent a judgment of low degree of continuity with the past.

Degree of Tension or Anxiety. The following instructions were given to the raters: "Rate the level of anxiety implicit in the interview material. This refers not only to manifestations of subjectively experienced anxiety but to the over-all level of tension or conflict inferred to be present in the personality structure of the individual." High ratings on this scale represent judgments of a high degree of anxiety.

The average of the three judges' ratings were taken as the best estimates on these two interview measures.[4]

Semantic differential measures. An adaptation of the semantic differential technique (Osgood, 1952) was used to measure the other two qualities characteristic of identity diffusion. This technique consists of asking *Ss* to describe their feelings about various concepts by rating them one at a time on a series of five-point scales.[5] The rating profiles of different concepts can then be compared as an indirect measure of the degree to which the *S* feels that these concepts are related to one another. *Ss* were asked to make semantic differential ratings on the 15 concepts of Table 1 during an initial group testing session. Four weeks later *Ss* again rated the five concepts marked by asterisks.

The first 14 "descriptive" concepts were selected as representing a broad range of possible important aspects of the *Ss*' self-conceptions. They were chosen to survey the major types of interpersonal behaviors. (For a discussion of the basis for selecting these concepts see Bronson, 1957). Since a person's identity is in large part the product of interpersonal experience, it is assumed that these 14 concepts are in a rough way representative of the universe of major possible self-conceptions.

TABLE 1

CONCEPTS RATED ON THE SEMANTIC DIFFERENTIAL

effectiveness[a]	bitterness
weakness	competence
kindness[a]	irritableness
forcefulness	sympathy
worthlessness	cooperativeness
rebellion	my "characteristic self" (yourself as you most often feel about yourself)[a]
tenderness	
anger[a]	
submissiveness[a]	

[a] Repeated ratings were obtained four weeks later on these five concepts.

Certainty of Self Conception. By comparing semantic differential profiles it is possible to measure the extent to which a *S*'s feelings about his "characteristic self" are similar to his feelings about each of the 14 modes of interpersonal behavior. It is assumed that close profile agreement means that the particular mode of behavior is an important aspect of the individual's sense of self. In making their semantic differential ratings of the 14 descriptive concepts the *Ss* were not oriented towards making self-judgments. The final concept—"my characteristic self"—was presented only after the completion of the descriptive concepts. Profile agreement therefore represents a measure of similarity of affective reaction to the concept of self and certain modes of interpersonal behavior.

The particular interest here, however, is not the *content* of the individual's identity—rather it is concerned with the degree to which the characteristic self can currently be identified with *any* dominant mode of behavior. Individuals in a state of identity diffusion should be uncertain in identifying themselves closely with any dominant interpersonal pattern. A measure of the degree to which the characteristic self can be identified with any interpersonal pattern is the extent of similarity between the profile of the "characteristic self" and that of the single descriptive concept most similar to it.

A ΣD^2 measure of profile agreement was used, thereby considering elevation as well as shape in the profile comparisons. To correct for a marked skewness in the distribution of these scores based on squared values, the square root of the ΣD^2 value was used as the measure of "certainty of self-conception." This single value thus represents the extent to which the *S* affectively equates his characteristic self with a particular interpersonal mode of behavior. A high score indicates a lack of agreement between the semantic differential rating of characteristic self and any of the ratings made by the *S* of the various modes of interpersonal behavior.

Temporal Stability of Self-rating. The second semantic differential measure is the constancy over time of

[3] My thanks to my wife, Wanda Bronson, and to Albert Shapiro, for aiding me in making these ratings.

[4] Intercorrelations among the three judges were: "continuity" ratings .56, .79, and .43; anxiety ratings, .55, .68, and .63. The reliabilities of these two scales computed by the intraclass correlation technique are .69 and .71, respectively.

[5] The semantic differential scales were: high–low, rough–smooth, empty–full, small–large, cold–hot, clear–hazy, fresh–stale, peaceful–ferocious, clean–dirty, sick–healthy, angular–rounded, soft–hard, wet–dry, tasty–distasteful, slow–fast. The majority of these scales are taken from Osgood. The criteria for scale selection was that the scales be remote from the domain of interpersonal concepts which were to be rated and that they be relatively free from evaluative connotations.

feelings about the characteristic self. This is expressed by the difference between the two "characteristic self" ratings spaced by a four week interval. Persons firm in their sense of self should show less affective variation in successive semantic differential ratings of self. The square root of the ΣD^2 value was again used to avoid a highly skewed distribution of the measure.

In order to determine whether *S*s showing test-retest inconsistency of "characteristic self" ratings were truly uncertain in their feelings about themselves, or were merely unable to use the semantic differential in a meaningful and consistent manner, a second measure of test-retest stability was obtained—a measure of constancy of ratings of the four descriptive concepts which were repeated in the second session. This measure is the sum of the test-retest profile differences for these four repeat concepts, again taking the square root to correct for skewness. If the measure of the temporal stability of self-rating truly reflects uncertainty about self, and not merely a general uncertainty in making any semantic differential rating, it should correlate with the other measures of identity diffusion while this latter measure reflecting general semantic differential rating fluctuation should not show such relationships.

RESULTS AND DISCUSSION

The product moment correlations among the four measures postulated as related to aspects of identity diffusion are presented in Table 2.

The measure of test-retest stability of the four repeated descriptive concepts correlates .33 with the measure Temporal Stability of Self-rating (Measure 4). However, it shows little correlation with the interview measures: .12 with the measure of continuity with the past, and .08 with the anxiety ratings. With this measure of test-retest stability eliminated through partial correlation, the measure of Temporal Stability of Self correlates with the latter two interview measures .52 and .47, respectively, showing no significant changes from the values reported in Table 2. Test-retest variability in ratings of "characteristic self," therefore appears to include variance other than that due to a general uncertainty in making semantic differential ratings.

TABLE 2

CORRELATIONS AMONG MEASURES RELATING TO ASPECTS OF IDENTITY DIFFUSION

	1 Continuity with Past	2 Freedom from Anxiety	3 Certainty of Self Conception
1 Continuity with the past (interview rating)	—		
2 Freedom from anxiety (interview rating)	.59**	—	
3 Certainty of self conception (semantic differential measure)	.71**	.47*	—
4 Temporal stability of self-rating (semantic differential measure)	.53*	.47*	.54**

Note:—N = 18 for the correlations involving the interview measures; N = 46 for the remaining value.

* $p < .05$, two-tailed test.

** $p < .01$, two-tailed test.

All relationships presented in Table 2 are significant in the predicted directions. *S*s judged from interview material to be lacking a stable foundation for their currently professed self-conceptions (Measure 1) show less certainty about the dominant characteristics of self (Measure 3), and manifest greater variability through time in their feelings about themselves (Measure 4). *S*s with high scores on these measures of identity diffusion also tend to be judged more anxious (Measure 2)[6]. Conversely, *S*s whose current identity impressed judges as an integrated expression of the personal past tended to identify themselves more closely with one or more dominant modes of interpersonal behavior, to be more temporally consistent in their feelings about themselves, and to be more free of anxiety.

The results of this study indicate that the construct of identity diffusion is a measurable parameter of personality development. This construct, reflecting variations in identity formulation, has promise as a genotypical variable interrelating diverse areas of adolescent behavior. Various characteristics of adolescence, such as the transient periods of social withdrawal or the seeking of psychological refuge in a commitment to a preformed social or religious ideology can be seen as symptomatic of, or as attempted solutions to, a condition of identity diffusion. In addition to its importance as an integrating construct in adolescent psychology, identity diffusion can also be seen as a significant concept in the study of adult personality differences. Different solutions to the problem of identity diffusion will lead to diverse personality characteristics in the adult. A premature identity closure could lead to a rigid denial of potentially

[6] It should be noted that the predicted correlation between Measures 1 and 2 could possibly represent a halo effect in making interview ratings.

creative personality components. On the other hand, a prolonged period of identity diffusion, which is conducive to a more complex manifest personality, may also interfere with the solid identity synthesis necessary for effective social action.

SUMMARY

Erikson has described late adolescence as the stage of "identity crisis" where individuals are often involved in a final redefinition of their sense of self. A frequent characteristic of persons at this stage of development is a state of "identity diffusion" in which the individual experiences uncertainty as to which of the manifold self-images derived from earlier self-conceptions and from projected ideal selves is most truly he.

In this study, four salient characteristics of identity diffusion are described: the current self-concept is less firmly rooted in earlier identifications; feelings about self show greater variability over short periods of time; the notion of self is less firmly conceived in terms of dominant interpersonal modes of behavior; and a state of heightened anxiety prevails. Measurement techniques are presented to measure these characteristics. Two measures are based on judgments from personal interviews with *Ss*, and two measures are based on semantic differential techniques. The four measures related to identity diffusion were tested on a sample of upper division college students, this being an age group which should show the greatest variation in degree of identity diffusion.

It was predicted that these measures of various aspects of identity diffusion would be intercorrelated. The data supported this prediction, each of the four measures being significantly correlated with the other three. *Ss* who would be considered low on identity diffusion impressed interview raters as having a stable sense of self rooted in previous history, and as being relatively free from anxiety. On semantic differential measures they were relatively certain about dominant personal characteristics, and showed temporal stability in their feelings about self. The opposite characteristics would describe *Ss* high on identity diffusion. Results lend support to the concept of identity diffusion as a significant variable descriptive of variations among persons of this age group.

REFERENCES

BRONSON, G. W. Variations in the structure of ego-identity. Unpublished doctoral dissertation, Univer. of California, Berkeley, 1957.

ERIKSON, ERIK H. The problem of ego identity. *J. Amer. Psychoanal. Ass.*, 1956, **4**, 56–121.

OSGOOD, C. E. The nature and measurement of meaning. *Psychol. Bull.*, 1952, **49**, 579–591.

(Received October 6, 1958)

ACCURACY OF BRIGHTNESS DISCRIMINATION AS MEASURED BY CONCURRENT VERBAL RESPONSES AND GSRs[1]

DON E. DULANY, JR. AND CHARLES W. ERIKSEN
University of Illinois

SUBLIMINAL perception experiments usually demonstrate that some response discriminates accurately below the threshold for a verbal response which the investigator identifies with awareness. Running through the several criticisms of this conclusion is the objection that the psychophysical assessment of the two responses is not comparable. Lack of comparability is especially objectionable, as Eriksen (1958) and Goldiamond (1958) have pointed out, when a phenomenal report procedure admitting several sources of extraneous error or control marks the threshold for awareness and a relatively uncontaminated accuracy judgment provides the measure of "subliminal accuracy." In the classical method of constant stimuli, the O is asked to give a phenomenal report of the presence or absence of a sensory experience on each trial. It has become clear that not only stimulus values but also the O's disposition to say Yes or No determine that report. For the typical undergraduate, wary of being thought hallucinated, the prudent response when unsure is No. This or any other kind of negative response bias can act to raise the phenomenal report threshold. Swets, Tanner, and Birdsall (1955) have shown that by placing "values" on true positives and true negatives and "costs" on false positives and false negatives in various combinations, the O's "operating level" or threshold varies accordingly. Other sources of nonsensory control—response category and serial effects—affect the phenomenal report to a significantly greater degree than the forced-choice verbal response (Blackwell, 1952, 1953; Goldiamond, 1958). Moreover, Eriksen (1958) has found striking individual differences in discrepancy between phenomenal report and forced-choice thresholds. The implication of this finding is that each O has his own criterion for the report of a sensory experience and that with the phenomenal report procedure, the threshold of awareness has substantially different meaning from O to O.

Eriksen (1956) and Howes (1954) have discussed an additional objection to comparing the accuracy of a discrete verbal response with the accuracy of a continuous GSR. When the O has only a limited number of discrete verbal responses with which to characterize a sensory experience, information may be lost that is not lost to the continuous GSR.

The object of the present study is to compare the discriminative accuracy of the GSR and verbal response when assessed by the same forced-choice psychophysical procedure.

METHOD

Subjects

The Os were eight male undergraduates paid for their service.

Design

Each O began the experiment with 10 one-hour sessions devoted to brightness discrimination. Ten hours, we had observed, put the O well along on a practice asymptote. Following these practice sessions, GSRs were conditioned to a light stimulus using electric shock as an US. Psychophysical functions for both verbal response and GRS were then obtained concurrently using a forced-choice procedure essentially like that of Blackwell (1953). With this procedure stimulus magnitudes are presented by the classical method of constant stimuli. On each trial, however, rather than a phenomenal report of the presence or absence of a sensory experience, O reports in which of several intervals the stimulus increment appeared; in this experiment, in which of two time intervals the light appeared. The time interval with the larger GSR for the trial was taken to be the GSR's "choice." In this way a discrete index of GSR association with a stimulus was obtained on every trial and the psychophysical measurement of GSR and verbal response became comparable. On the assumption that saying First or Second, as well as presence or absence of the stimulus and greater or lesser GSR, dichotomously indexes a continuous variable, tetrachoric correlation was computed as a measure of the association of the stimulus with each kind of response.

Setting and Apparatus

The O was seated in a 6′ × 4′ × 6′ light proof room contained within the main experimental room. Four feet in front of his face was a circular field of light 6″ in diameter with a homogeneous illumination level of 16.40

[1] This research was supported by National Science Foundation Grant 4461 (D. E. Dulany) and Mental Health Grant M-1206, National Institutes of Health, Public Health Service (C. W. Eriksen).

foot-candles. The test stimuli were presented directly above a fixation point in the center of this field.

The light source was a 300-watt G.E. projection bulb encased in a metal box and cooled by a blower. On one side of the box, an aperture permitted light to pass into a system of lenses, through a light tunnel and then illuminate the circular field presented to O. Through a second aperture at right angles to the first, the light could pass through a 1/64" hole, then through a circular wedge and by a system of prisms and mirrors be reflected through the light tunnel to the position of the test patch on the circular field. Each test stimulus was .6" × .2" in area and could be varied in intensity by positioning of the circular wedge. A Stoelting stimulus timer controlled a buzzer designating the stimulus intervals and also a shutter which in turn controlled the onset and duration of the test stimulus.

GSRs were recorded with the Stoelting Deceptograph modified so that direct reading of O's basal resistance level could be obtained. Contact was made by sponge electrodes soaked in saline solution and fastened to the sole of O's foot.

Electric shock was produced by a Harvard inductorium set for a faradic stimulation and supplied by a 3-volt DC source.

Procedure

Practice Sessions. At practice as well as later experimental sessions, O was allowed a 10-min. adaptation period before psychophysical judgments began. Four Os were trained by the phenomenal report method during the practice sessions, the other four by the forced-choice method and phenomenal report method alternately. Practice effects of the two procedures upon the forced-choice discrimination had been found not to be appreciably different.

In the phenomenal report procedure O was instructed that he was to say Yes when he observed the test light and "No" when he did not. Test lights were presented for 3 sec., the interval marked by a muffled buzzer that sounded throughout. Five seconds elapsed between presentations. Each practice session consisted of a total of 240 judgments composed of 24 judgments of each of nine different stimulus intensities plus a blank interval. The frequency of Yes responses to the blank interval defined O's false alarm rate.

In the forced-choice procedure O was instructed that the test light would occur in one of two 3-sec. intervals, each signaled by a buzzer. He was to report at the end of the trial whether the light appeared in the first or second interval. Three sec. elapsed between intervals with a 5-sec. pause between the end of the second interval and the beginning of the next trial. Twenty-four presentations of each of nine stimulus intensities comprised a practice session.

For both procedures, order of presentation of the nine stimulus intensities and the blank interval was randomized, with the restriction that each stimulus and the blank occur twice within each block of 20 trials. In neither procedure was a doubtful category of response permitted. Practice sessions were run on consecutive days at approximately the same time of day for each O with the exception that Os were not available for Sunday running.

GSR Conditioning and Psychophysical Assessment. GSR conditioning began for each O on the day following his tenth practice session. On each trial the buzzer sounded and the light appeared at an intensity that O had reported Yes between 90 and 100% of the time during practice in the phenomenal report procedure. The first seven presentations of the light were reinforced with a moderate electric shock to the forearm. After the seventh trial, partial reinforcement was gradually adopted until by the twentieth trial, O was on a 30% reinforcement schedule. Blank intervals with the buzzer alone were inserted to extinguish GRS to the buzzer and insure a discriminated response to the light. Conditioning trials continued until an unmistakable conditioned response occurred on nonreinforced presentations of the light, and little or no response occurred during the blank interval with the buzzer alone. An effort was made to condition to a maximum, halting before adaptation began.

Immediately after conditioning, psychophysical functions for concurrent forced-choice verbal response and GSR were obtained. The forced-choice procedure of this session was identical with that of the practice sessions with two exceptions: On each trial, maximum magnitude of GSR to each of the two stimulus intervals was recorded; and instead of the blank stimulus interval, a reinforced presentation of the light stimulus at conditioning intensity occurred twice among every 20 judgments. Since the GSR values to be compared occurred only a few seconds apart, drift in the GSR baseline was not a problem.

To insure adequate GSR conditioning Os submitted to as many as three sequences of conditioning and psychophysical assessment. The criterion adopted was a significant ($p < .05$) positive tetrachoric correlation between presence of a greater GSR and presence of the stimulus computed over the nine stimulus intensities (108 judgments). Four Os met this criterion at the second session and four more at the third session. Another three Os did not meet the criterion by the third session and were dropped from the experiment.

A correction was made for possible first or second position biases in GSR. Maximum GSR occurring four seconds after the beginning of each stimulus interval was converted to micromohs of conductance change. The average GSRs at first and second interval for successive series of 20 judgments were plotted for each O. Curves were fitted to these two sets of points and the difference between the two curves for a judgment became the bias correction for that judgment. This correction was probably unnecessary since analysis of the data without it yielded almost identical results.

RESULTS AND DISCUSSION

For Os A through D the psychophysical functions for GSR and forced-choice verbal response (see Fig. 1)[2] are those obtained during

[2] Due to apparatus difficulty, data were usable at only six stimulus values for the first two Os (A and B) and at eight stimulus values for the second two Os (C and D). The effect of this loss of data was to restrict the range of accuracy such that for the first three Os, accuracy did not excel 93% for either response. This restriction does not seem critical.

the second experimental session; for the remaining Os, those obtained during the third experimental session. Phenomenal report curves come from each O's final practice session and are included for rough comparison. We have chosen to present individual psychophysical functions because any method of arriving at composite curves is likely to introduce artifacts critical for this experimental problem.

Our first concern is to know whether either GSR or verbal response shows superior discriminative accuracy when assessed under comparable forced-choice conditions. For seven Os analyzed individually, the r_t of forced-choice verbal response and stimulus computed across all stimulus values was greater than that of the GSR and the stimulus. To evaluate the significance of this difference, r_ts for both kinds of response were converted to Fisher's z and averaged over Os. The resulting r_ts for the forced-choice verbal response are .647 and for the GSR .423. The t for this difference is 2.73, which with 7 df is significant beyond the .05 level. Over a range of stimulus values producing from chance to 99% accuracy the verbal response was the better discriminator.

Since the more common contention is that it is the threshold range that is the area of

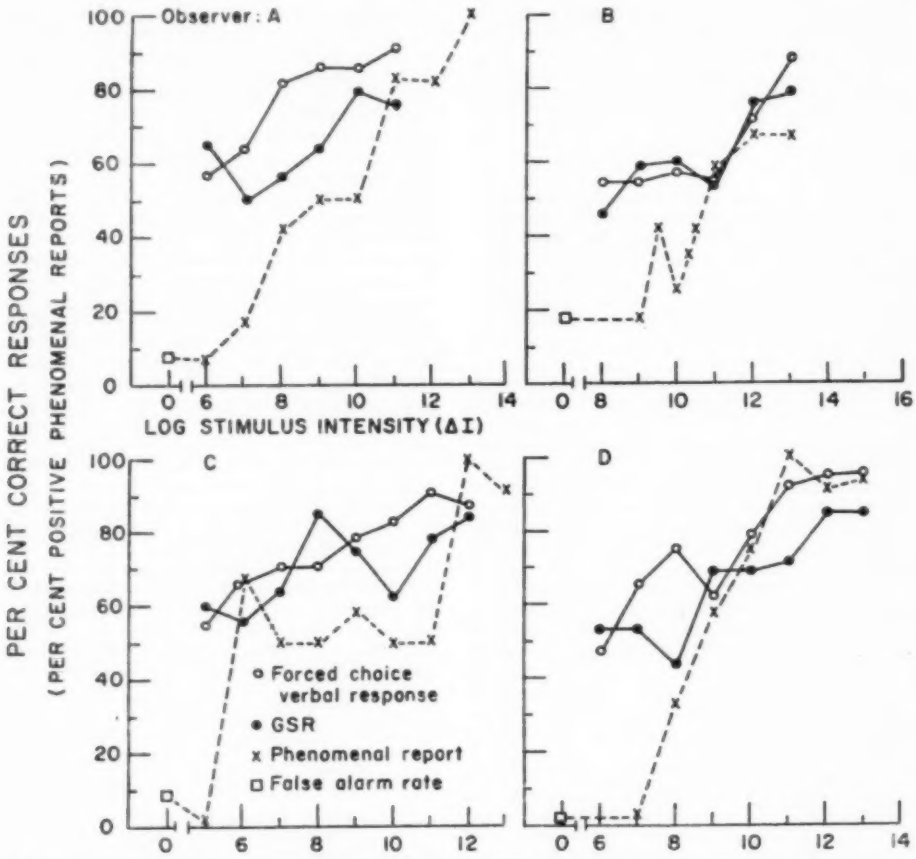

FIG. 1. PERCENTAGE OF CORRECT VERBAL RESPONSES AND GSRs, AND PERCENTAGE OF POSITIVE PHENOMENAL REPORTS AS A FUNCTION OF LOG STIMULUS INTENSITY
On this scale, 2 = −.72 log foot-candles and each scale unit equals .08 log foot-candles.

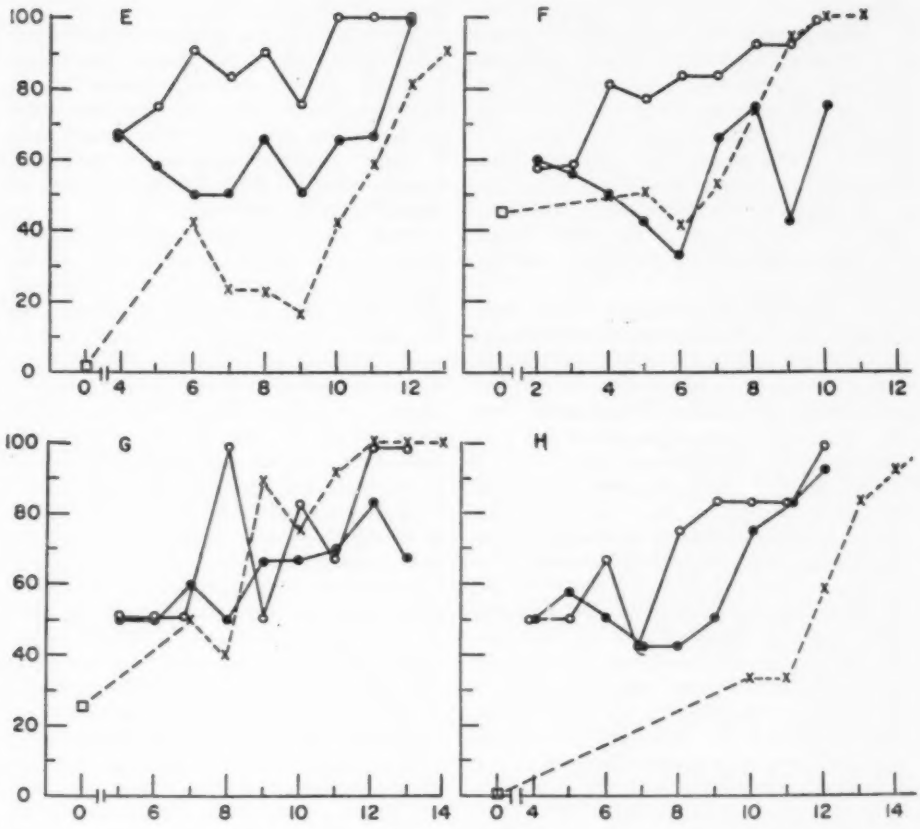

FIG. 1. *Continued*

greater GSR sensitivity, this analysis was repeated over the three lowest stimulus values. For six Os the forced-choice verbal response was more highly correlated with the stimulus than was the GSR. For one O, the GSR correlation with the stimulus was higher, and for the other neither correlation was appreciable (−.06 and .01). Averaging by means of Fisher's transformation yields a mean r_t of .371 for the forced-choice verbal response as compared with a mean r_t of .309 for the GSR. A t of 1.51 for this comparison falls short of the .05 level. There is no evidence of superior discriminative accuracy for either response in this range.

We have chosen not to speak of thresholds and their comparisons. The very concept of threshold has been subjected to critical analysis (Tanner & Swets, 1954; Swets, Tanner, & Birdsall, 1955) and may be dispensable. Any empirical concept of the threshold is statistical and most are, in some degree, arbitrary. At least some of the confusion surrounding the concept of subception can be laid to inconsistent conceptions of what constitutes a threshold. Instead, individual curves are presented for visual inspection. If there is a single number for each of these functions that can be most meaningfully compared, it should be the inflection point of each curve where it begins to depart from a chance baseline. This ought to be the point at which the response begins to show discriminative accuracy. For the forced-choice verbal response and the GSR, the

chance line is 50%; for the phenomenal report, the usual estimate for the chance line is the false alarm rate. Although all three inflection points agree rather well, there is some tendency for the inflection to occur earlier for the forced-choice verbal response than for either of the other responses (Os A, D, G). Most important, in seven cases the inflection point of the curve for the forced-choice verbal response is below that of the GSR. These comparisons are entirely consistent with the tetrachoric analysis.

The substance of this finding is that when discrete indices of both the verbal response and the GSR are compared, the GSR is no better, and considering a wide range of stimuli presented, is the inferior discriminator. Since the verbal response as manifested is normally discrete it is interesting in any case to compare similar measures. But the generality of the finding is as good as the assumption that verbal response tendency—excitatory potential, a priori probability—is continuously distributed.[3] Otherwise the GSR could be said to suffer from an unfair restriction placed upon it. In the course of ordinary experience it could be the very continuity of whatever the GSR indexes that makes it a better discriminator than the discrete verbal response. The assumption of continuity of response strength, however, is common for nonlinguistic instrumental responses and well supported. It is also an assumption made by Skinner (1957) and Osgood (1957) in their treatments of verbal behavior. The specific assumption of continuity of verbal response tendency in psychophysical measurement is supported by the experiments of Bricker and Chapanis (1953), Murdock (1954), and Tanner, Swets, and Birdsall (1955). These experiments show that when the initial response is incorrect, a subsequent guess or ranking of possible responses agrees with the stimulus with greater than chance expectancy. When the correct response is not raised to evocation strength, the stimulus still increases its probability of occurrence as revealed by subsequent measures. With this assumption basic to his statistical theory of subception, Howes (1954) has been able to deduce the relevant experimental facts. At each trial of the present experiment, GSR strength in micromohs change for the two stimulus intervals may show any ratio marked by the extremes of 0 to 1 and 1 to 0. It seems entirely reasonable that verbal response tendency expressed in p values is distributed in the same way.

There is an important question whether these training procedures permitted both the GSR and the verbal response to attain their greatest accuracy. These Os had experienced 10 hours of verbal judgment and were in all cases clearly at an asymptote. Training with the electric shock did not, and for practical reasons, could not be extended through 10 hours. What the GSR accuracy would have been after 10 hours of training we do not know. With the possibility of adaptation or some general disorganization from long periods of shock, there was concern that accuracy would eventually decline. Conditioning was halted at what appeared to be an asymptote, but because of the instability of the GSR a "maximum of accuracy" is difficult to recognize and a true asymptote may not exist. This difficulty obviously limits the usefulness of the GSR as an index of autonomic or emotional sensitivity. Our conditioning procedures have, however, been more extensive than is typical of experiments in this area. We do have clear evidence that conditioning occurred and was maintained in the significant r_{tS} for GSR and stimulus and in the psychophysical functions for the GSR, which are in fact curves of generalization from the point of conditioning, a clear and visible stimulus light. Moreover, seven of eight Os met a criterion of 80% correct GSR "judgments" and the other exceeds 75% accuracy during the assessment series. With 1 in every 10 occurrences of the stimulus associated with shock during psychophysical assessment, it was unlikely that the GSR would suffer appreciable extinction, and apparently it did not.

There are, of course, features of the measurement of GSR and verbal response that are not identical here but they seem inherent in the measurement of these responses. Any error in the GSR measuring instrument subtracts from the correlation of GSR and stimulus.

[3] When the accuracy of the continuous GSR is compared with the accuracy of some discrete verbal response taken to be an index of sensory awareness (e.g., Lazarus & McCleary, 1951), the measures are not comparable if either the sensory or verbal response dimension is continuous.

And similarly sources of unidentified error for the verbal response will reduce the stimulus-verbal response association. As Eriksen (1956) has shown, there is an appreciable degree of independence between the errors in verbal responses and GSRs. Variables that will influence the accuracy of verbal responses will not concurrently affect the GSR accuracy, at least not to the same extent. Moreover, even though extraneous sources of error in the verbal response may be minimized with the forced-choice technique, many for the GSR—muscle movement, transient fears and discomforts—cannot so easily be ruled out. It is these latter that seem to place the main limits on the accuracy of the GSR.

The subception hypothesis is usually taken to mean that one can, with some response or another, discriminate stimulus values too weak to produce awareness. This hypothesis has not been convincingly tested and it is not tested here. The forced-choice verbal response is not what most would intuitively identify with awareness. And, in fact, numerous studies have reported accuracy excelling chance of some kind of forced-choice response at stimulus intensities that do not provoke the O to say "I can see it" or its equivalent (see Adams, 1957; and Goldiamond, 1958 for reviews). The current objection to these studies (Goldiamond, 1958) is a reassertion of the old argument that reports of private experience are suspect, offered with new documentation that phenomenal reports are under the control of nonsensory variables (Blackwell: 1952, 1953; Tanner & Swets, 1954; Swets, Tanner, & Birdsall, 1955). Until more acceptable techniques for manipulating or indexing awareness are available, it seems more sensible to ask what response systems show greatest discriminative accuracy under comparable conditions of measurement.

SUMMARY

In many subliminal perception experiments some response shows discriminative accuracy at stimulus values below the threshold of awareness determined by classical psychophysical techniques. The procedure is objectionable because the threshold measure admits extraneous variance that the measure of "subliminal accuracy" does not. The measures are not comparable for another reason when one response is the GSR and the other a verbal response because the former is continuous and the latter, as manifested, is discrete. This experiment compares the discriminative accuracy of GSR and verbal response when assessed by the same forced-choice psychophysical technique, a procedure that obtains a discrete index of both responses.

1. Over a range of stimulus values producing from chance (50%) to approximately 99% accuracy of judgment, the verbal response showed significantly higher correlation with the stimulus than did the GSR.

2. Within the range of stimulus values producing from chance to approximately 70% accuracy of judgment, neither response was significantly more accurate than the other.

REFERENCES

ADAMS, J. K. Laboratory studies of behavior without awareness. *Psychol. Bull.*, 1957, **54**, 383–405.

BLACKWELL, H. R. The influence of data collecting procedures upon psychophysical measurement of two sensory functions. *J. exp. Psychol.*, 1952, **44**, 306–315.

BLACKWELL, H. R. Psychophysical thresholds: Experimental studies of methods of measurement. *Engng. Res. Bull.*, Univer. of Michigan, 1953, **36**, 1–227.

BRICKER, P. D., & CHAPANIS, A. Do incorrectly perceived stimuli convey some information? *Psychol. Rev.*, 1953, **60**, 181–188.

ERIKSEN, C. W. Subception: Fact or artifact? *Psychol. Rev.*, 1956, **63**, 74–80.

ERIKSEN, C. W. Unconscious processes. In *Nebraska symposium on motivation*. Lincoln: Univer. Nebraska Press, 1958.

GOLDIAMOND, I. Indicators of perception: I. Subliminal perception, subception, unconscious perception: An analysis in terms of psychophysical indicator methodology. *Psychol. Bull.*, 1958, **55**, 373–412.

HOWES, D. H. A statistical theory of the phenomenon of subception. *Psychol. Rev.*, 1954, **61**, 98–110.

LAZARUS, R. S., & MCCLEARY, R. A. Autonomic discrimination without awareness: A study of subception. *Psychol. Rev.* 1951, **58**, 113–122.

MURDOCK, B. B., Jr. Perceptual defense and threshold measurements. *J. Pers.*, 1954, **22**, 565–571.

OSGOOD, C. Motivational dynamics of language behavior. In *Nebraska symposium on motivation*. Lincoln: Univer. Nebraska press, 1958.

SKINNER, B. F. *Verbal behavior*. New York: Appleton-Century-Crofts, 1957.

SWETS, J. A., Tanner, W. P., Jr., & Birdsall, T. G. The evidence for a decision-making theory of visual detection. *Tech. Rep. No. 40, Electronic Defense Group*. Ann Arbor: Univer. of Michigan, 1955

TANNER, W. P., JR., & SWETS, J. A. A decision-making theory of visual detection. *Psychol. Rev.*, 1954, **61**, 401–409.

(Early publication received April 30, 1959)

CRITIQUE AND NOTES

A SEARCH FOR CORRELATES OF INTRA-INDIVIDUAL RESPONSE VARIABILITY[1]

WALTER J. RAINE
Menninger Foundation
AND JOHN R. HILLS
Regents of the University System of Georgia

IN A recently reported study (Raine & Hills, 1959) the authors have provided evidence that two of Fiske and Rice's (1955) concepts of intra-individual response variability operate very similarly when reliable scores reflecting these concepts are obtained from answers to sentence-completion items. In the present study (Raine & Hills, 1958), we correlated both of our variability scores with a wide variety of measures, 39 in number, including all 25 of the Personality Research Inventory (PRI) scores and three PRI response sets to answer Yes, No, or ? regardless of item content (Saunders, 1955), 4 measures of motivation (Hills: 1955, 1958), law school grades and Law School Admission Test scores (LSAT), as well as several "flexibility" measures (American Council of Education, 1953; Frederiksen; 1958a, 1958b; Helmstadter, 1957; Rokeach, 1956; Stern, Stein, & Bloom, 1956). The *S*s were male entering law students at an eastern university. Attendance at the testing session was obligatory; the students were assured that the individual results were confidential. The final *N* for this study was 43. The sentence completion test was scored as it was scored by Fiske[2]; however, our results are presented in terms of variability, rather than similarity.

RESULTS

Of the correlations between the 39 variables and the two measures of intra-individual response variability[3], only two are significantly different from zero. We cannot say much on the affirmative side concerning the correlates of intra-individual response variability. When this many statistically dependent significance tests on correlations are computed, the only significant correlations in our data could have occurred even though the null hypothesis were true for the entire set of hypotheses. Our evidence does not support Fiske's (1957) finding that variability on free response tests is positively related to PRI Tolerance of Ambiguity (called Tolerance for Complexity by Fiske) and is negatively related to PRI Gregariousness (Fiske's Scale 11).

We are left with an enigma. Variability in responding to the same or similar items is considered by testers to be error variance. But the tendency for an individual to vary is as highly reliable as scores from our better psychological tests. The problem is that if intra-individual variability is true variance it seems odd that we cannot find any correlates of it. It is entirely possible that variability scores, along with response set scores, are measuring aspects of personality that do not appear in self-report scores. The search for personality-measure correlates of variability should, perhaps, be directed toward other measures of style rather than content (Hills, 1957; Jackson & Messick, 1958).

REFERENCES

AMERICAN COUNCIL ON EDUCATION, Committee on Measurement and Evaluation of the Cooperative Study of Evaluation in General Education. *Instructor's manual for the Inventory of Beliefs*. Washington: Author, 1953.

FISKE, D. W. The constraints on intra-individual variability in test responses. *Educ. psychol. Measmt.*, 1957, **17**, 317–337.

FISKE, D. W., & RICE, LAURA. Intra-individual re-

[1] This research is part of a larger study conducted by Educational Testing Service with the cooperation of several member schools of the Law School Admission Test Policy Committee. The project was under the direction of John R. Hills. The authors were employees of Educational Testing Service during the study. The help of the law schools participating in the study is greatly appreciated.

[2] D. W. Fiske transmitted his criteria for scoring the sentence completion test to the authors in a personal communication. The authors are most appreciative of D. W. Fiske's interest in and help with their work on intra-individual response variability.

[3] A two-page table giving the means, standard deviations, and reliabilities of the 39 scores and the correlations between these measures and the intra-individual response variability scores has been deposited with the American Documentation Institute. Order Document No. 6037, from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington, 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make check payable to Chief, Photoduplication Service, Library of Congress.

sponse variability. *Psychol. Bull.*, 1955, **52**, 217–250.

FREDERIKSEN, N. O. *A study of tests of report writing ability.* (*ONR Tech. Rep.*) Princeton, N. J.: Educational Testing Service, 1958. (a)

FREDERIKSEN, N. O., & MESSICK, S. *Response set as a personality variable.* (*ONR Tech. Rep.*) Princeton, N. J.: Educational Testing Service, 1958. (b)

HELMSTADTER, G. C. Procedures for obtaining separate set and content components of a test score. *Psychometrika*, 1957, **22**, 381–394.

HILLS, J. R. The measurement of levels of aspiration. *J. soc. Psychol.*, 1955, **41**, 221–229.

HILLS, J. R. Quasi-cognitive tests of personality. *Amer. Psychologist*, 1957, **12**, 385. (Abstract)

HILLS, J. R. Needs for achievement, aspirations, and college criteria. *J. educ. Psychol.*, 1958, **49**, 156–161.

JACKSON, D. N., & MESSICK, S. Content and style in personality assessment. *Psychol. Bull.*, 1958, **55**, 243–252.

RAINE, W. J., & HILLS, J. R. Correlates of intra-individual response variability. (Res. Bull. 58–9) Princeton, N. J.: Educational Testing Service, 1958.

RAINE, W. J., & HILLS, J. R. Measuring intra-individual variability within one testing. *J. abnorm. soc. Psychol.*, 1959, **58**, 264–266.

ROKEACH, M. Political and religious dogmatism: An alternative to the authoritarian personality. *Psychol. Monogr.*, 1956, **70**(18, Whole No. 425).

SAUNDERS, D. R. Some preliminary interpretive material for the PRI. (Res. Memo. 55–15) Princeton, N. J.: Educational Testing Service, 1955.

STERN, G. G., STEIN, M. I., & BLOOM, B. F. *Methods in personality assessment.* Glencoe, Ill.: Free Press, 1956.

(Received July 21, 1958)

THE DEVELOPMENT OF SHARING BEHAVIOR

BRITOMAR J. HANDLON AND PATRICIA GROSS

Mount Holyoke College

In a study on the moral judgments of children in Istanbul, Uğurel-Şemin (1952) asked children between the ages of 4 and 16 years to divide an unequal though varying number of nuts with another child, presumably of comparable age. The sharing decision was made verbally to an adult in the absence of the partner who was then brought in and the sharing act carried out. She found that at four to six years 67% of the children kept a greater share for themselves and gave a lesser share to the partner (were "selfish") and that with increasing age an increasing proportion of the children did not keep the greater share but either gave it to the partner (were "generous") or made an equal division and disposed of the odd nut in various ingenious ways. Children six to eight more frequently gave the partner the larger share (were "generous") and older children more frequently made an equal division with two exceptions at ages 12–13 and 14–15 where they were more frequently "generous." It is the thinking of the present writers that the alternation between equal sharing and "generosity" may have been a result of the specific instructions which apparently varied from child to child and which seemed to focus attention to a greater or lesser degree on the indivisible nut, so that to some children the situation may have been presented as a conundrum to be solved, while to others the situation may have been presented as an ethical problem.

Uğurel-Şemen also found that children from small families (one or two children) were "selfish" more often than children from larger families, that sex was not an important variable, that poor children were "generous" or made an equal division more frequently than the rich, and that the middle-class children were more frequently "selfish" than either rich or poor.

The present study is pertinent only to Uğurel-Şemin's findings concerning sharing behavior. No attempt is made to evaluate her conclusions concerning the development of generosity, equalitarianism, and moral judgments, conclusions which were based primarily on verbal behavior elicited at the time of the sharing decision.

The effect of age, sex, and only or other-than-only status on sharing behavior was studied in the present experiment. An adult was always present, and the sharing act was carried out in the absence of the partner who left the room after the two children had cooperated to earn pennies or seals as the case might be. The effect of the presence of an adult was not studied but an earlier unpublished study by Atwood and Sengstaken (1951) suggests that children are more likely to be "generous" in the presence of an adult.

METHOD

Apparatus

The apparatus used in this experiment is almost identical with that developed by Sollenberger for

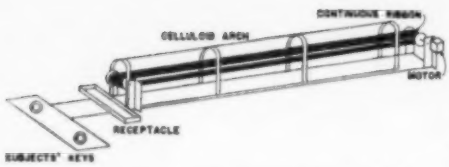

FIG. 1. DIAGRAM OF DELIVERY APPARATUS

an earlier unpublished study (Shillady & Ellis, 1948).

It consists of an unpainted wooden board 9 cm. wide and 105 cm. long. Two sets of metal strips are nailed 10 cm. from each end of the board, to which are attached belt pulleys, each with a circumference of 13.2 cm. The belt used is a brightly colored ribbon. This ribbon is 8 cm. wide and 85 cm. long when measured from one pulley to the other. It is kept taut by a spring screwed into the board and attached to a metal strip at the *S*'s end of the board. A sheet of celluloid is arched over the belt, nailed to the board, and braced by three strips of metal. The arch thus formed is open at both ends. On the left side of one end of the board is a telechron motor which allows the pulley to make one revolution per second. Two telegraph keys are wired in series so that when they are operated simultaneously, the motor is activated and the belt revolves at a speed of 13.2 cm. per second. It takes 6.4 seconds for an object placed on one end of the "conveyor belt" to reach the other end.

Subjects

The *S*s in the first experiment, called Group A, were 10 preschool (nursery and kindergarten) children and 11 from the fourth, fifth, and sixth grades of a local public school. At each grade level there was an even number of only children and children with siblings, except in one case when an *S* became ill and did not arrive for a session.[1]

The *S*s in the second experiment, Group B, consisted of eight preschool (nursery and kindergarten) children and 14 children from the fourth, fifth, and sixth grades from the local public school. There were no only children in this group, since all available only children had been used in the first experiment.

Procedure

In the experimental situation every child in a given grade was paired with every other child of the same sex in that grade who had been selected for that experiment. The pairs were introduced to the experimental situation in a random order.

Two *S*s sat at the end of the apparatus with the keys. They were instructed thus:

> See this tunnel with the ribbon in it. Well, you can make the ribbon move. Do you see the two keys? Each of you put a finger on one of them. Now if you press down at the same time the ribbon will move. Try it. O.K.? Now I am going to put some pennies (or seals in the case of the preschool children) on the other end of the ribbon. Your job is to make them come to you. Do not put your hand in the tunnel, but let the pennies (seals) fall off when they get to your end.

A box was placed to catch the pennies and the *S*s' questions were answered in a noncomittal fashion. When the pennies or seals reached the children, the *E* asked one of the children to leave the room and sit in a chair outside the door while she picked up the pennies (or seals) and laid them flat on her hand. The following instructions were given to the child remaining with her:

> Here are the pennies (seals) that you and (Johnny) earned. Since the two of you earned them you may keep them. Put the ones that are for Johnny in this pile, and put the ones that are for you in your pocket.

After the child made the division the supply of pennies or seals was brought up to five again and the child outside the room was given the same instructions. The children were alternated as much as possible in going out and staying in the room. At the end of an experimental session each child was given enough pennies or seals to make his total equal to that of the child who had taken the most. In addition, the elementary school children received a quarter for their help.[2]

RESULTS

Table 1 displays the results for Group A, Group B, the groups combined, only children and others, and boys and girls.

Group A. Looking at Group A first it can be seen that, with one reversal between nursery and kindergarten, as age increases the mean number of pennies or seals kept by the *S* for himself decreases. The between grades variance was assessed using

[1] The number of available only children in each grade was determined and then each of these was matched with a child of the same grade and sex who had one or more siblings.

[2] The preschool children were members of a nursery and kindergarten run as a laboratory school by Mount Holyoke College. The elementary school children were recruited from the local public school from among those children who volunteered for "child studies at the college" and whose parents had given them blanket permission to participate in such studies. It has been customary to give these children $0.25 for participation in each study.

TABLE 1
MEAN PERCENTAGES OF PENNIES OR SEALS RETAINED BY GROUPS BY GRADES

	N	*K*	4th	5th	6th	All Grades
Group A	68	72	50	43	40	56
Group B	78	72	66.5	43	40	58.5
Groups A and B	72	72	58	43	40	57
Only Children (Group A only)	63.5	60	53	40	40	53
Children other than only (Group A only)	70	83.5	46.5	45	40	59
Boys (Groups A and B)	72	72	—	42	40	55
Girls (Groups A and B)	72	72	58	43	40	60

the Kruskal-Wallis (Siegel, 1956) one-way analysis of variance by ranks. An H (interpreted as χ^2 with $df = 4$) of 15.5 was obtained. This is significant at better than the .01 level.

Since the children in different grades received a varying number of trials,[3] the percentage of pennies or seals kept by each child was first calculated and then these percentages were averaged for boys and girls separately and for only children and others separately. The boys kept an average of 56% of the rewards, and the girls 57%; a difference obviously not significant. The only children kept 53% of the rewards and the other-than-only children kept 59%. A t test showed this difference not significant.

Differences between only children and others are present at some grades, but since the over-all difference between only and other-than-only children is not significant these are probably chance variations, particularly since the direction of the differences is in the opposite direction from that which was predicted from Uğurel-Şemin's findings.

Group B. The age-trend results found in Group A were checked with Group B. None of the *Ss* in Group B was an only child (all available only children had been used in Group A), but otherwise the experimental situation remained identical.

From Table 1 it can be seen that the age trend found in Group A is repeated in Group B with no reversals in the means at any point. The between-grades variance was assessed using the Kruskal-Wallis one-way analysis of variance by ranks ($H = 12$, $df = 4$, $p < .02$). Thus, age as a significant factor was again demonstrated.

The difference between boys and girls was in the same direction as in Group A (boys kept 54% and girls kept 63%), but this difference was not significant when assessed by a t test.

Since both samples were drawn in the same manner from the same available population, it seemed reasonable to assume that they were truly samples from the same population. The groups were therefore combined for further analysis.

Groups combined. From Table 1 the age trend for the groups combined can be seen clearly, and there are no reversals in the means. As expected, a significant age factor in the groups combined was found using Kruskal-Wallis one-way analysis of variance by ranks ($H = 22.82$, $df = 4$, $p < .001$).

It is noteworthy that the difference is most marked between preschool and elementary grades.[4] A t test of the preschool–elementary difference gave a $p < .001$. It can therefore be seen that "generous" behavior (as assessed by sharing with an absent partner in the presence of an adult) becomes established as a norm in the elementary grades.

The transition between keeping the greater share and keeping the lesser share, when the child is forced to divide an unevenly divisible number of pennies or seals in the presence of an adult, seems to come between the fourth and fifth grades. In the fourth grade, six *Ss* kept more than 50% and only two kept less, whereas in the fifth grade, only two of the nine *Ss* kept more than an average of 40%. In the sixth grade, all *Ss* kept 40% on every trial. That is, they kept two pennies for themselves and left three for their partner every time.

SUMMARY AND CONCLUSIONS

The purpose of this experiment was to study sharing behavior in children as it is affected by age, sex, and only or other-than-only status. The experimental situation was that of sharing, with a momentarily absent partner, and in the presence of an adult, an unevenly divisible reward earned by performing a cooperative task. Forty-three New England preschool and grade school children were

[3] It will be recalled that each *S* was paired with every other *S* of the same sex and grade, but that the number of *Ss* in any grade was determined by the number of only children available in that grade.

[4] It has to be assumed of course that the sample from the preschool is not from a different population than the sample from the public school. That this assumption can be made is asserted by the following logic: The two samples are from the same geographical area and many of the available grade school children are relatives of those in the preschool. Also it is probable the same factors were operating in those parents who sent their children to a laboratory school and those who gave their grade school children permission to participate in child studies at the college. In fact, many of our *Ss* were earlier "turned down" at the laboratory school because of enrollment limitations. For these reasons it is felt that the comparison of the two groups compares age alone and not socioeconomic or other factors.

studied with the following results:

1. Age was a significant variable in sharing behavior. Giving the partner the greater share of an unequally divisible number of objects increased with age.

2. The highest degree of keeping a greater share for the self ("selfishness") occurred in the preschool group.

3. The transition between keeping the greater share ("selfishness") and keeping the lesser share ("generosity") occurred between the fourth and fifth grades and was complete by the sixth grade.

4. Sharing behavior was not influenced by the sex of the child.

5. The fact that a child was an only child or had siblings did not affect the sharing behavior of the children studied.

Our study of sharing behavior in New England children confirmed most of Uğurel-Şemin's findings on children in Istanbul. Two of her findings were not confirmed: (*a*) Her finding that size of family is an important variable was not confirmed. This result may reflect a cultural difference or it may be due to the small number of only children represented in our study. (*b*) Her finding that giving away the larger share ("generosity") was not a smoothly increasing function of age was not confirmed. It is speculated that her finding is an artifact of the nonstandard instructions to the subject and the use of the ambiguous equal-division category.

REFERENCES

ATWOOD, SALLY, & SENGSTAKEN, BARBARA. The differences in the frequency and variability of sharing responses between three- and five-year-old children. Unpublished manuscript, Mount Holyoke College Psych. Library, 1951.

SIEGEL, S. *Nonparametric statistics*. New York: McGraw-Hill, 1956.

SHILLADY, JANET, & ELLIS, EMILIE. A study of cooperation in three different age groups in a potentially competitive situation. Unpublished manuscript, Mount Holyoke College Psych. Library, 1948.

UĞUREL-ŞEMIN, REFIA. Moral behavior and moral judgments of children. *J. abnorm. soc. Psychol.*, 1952, **47**, 463–474.

(Received July 2, 1958)

SELF-DISCLOSURE AND OTHER-CATHEXIS

SIDNEY M. JOURARD

University of Florida

THE amount of personal information that one person is willing to disclose to another appears to be an index of the "closeness" of the relationship, and of the affection, love, or trust that prevails between the two people. In more general terms, self-disclosure and cathexis for the other person may be said to be correlated. Evidence to support this proposition stems from both clinical observation and systematic research. Thus, psychotherapists have long noted that when a patient feels warmth, trust, and confidence in his therapist, he discloses himself more freely and fully than when he perceives the therapist as hostile, punitive, or when he dislikes the therapist. More direct evidence of the relationship between other-cathexis and self-disclosure is provided by the work of Jourard and Lasakow (1958) who found significant correlations between questionnaire measures of "mother-cathexis" and "father-cathexis," and the amount that subjects (*Ss*) had disclosed to their parents.

The present study is an extension of the inquiry into the relationship between self-disclosure and other-cathexis. In addition, inquiry was directed into (*a*) the relationship between a person's cathexis for the other and the amount of personal information that he knows (from past disclosures) about the other person, and (*b*) the extent to which disclosure to others and knowledge about others involve reciprocal relations between people. The latter aim may be regarded as an indirect test of the hypothesis that one of the necessary conditions for promoting self-disclosure in another is to volunteer it oneself. On this assumption, a person who discloses much to another may be expected to receive much disclosure from him; and a person who knows a great deal about another should be relatively well-known by him. Indifference or antipathy between two persons may be expected to produce the consequences of low disclosure to one another, and little knowledge about one another as persons.

The hypotheses of the study were therefore as follows:

1. The amount of personal information that a person has disclosed to each of a group of coworkers covaries with his cathexis for each fellow group-member.

TABLE 1
THE SELF-DISCLOSURE QUESTIONNAIRE

1. When you were in college, did you participate in any extra-curricular activities, e.g., clubs, dramatics, sports, etc.?
2. In what town were you born?
3. Have you ever gone steady, or been engaged? If yes, what was his (their) name(s)?
4. Do you have any brothers or sisters? How many of each?
5. Do you have a main hobby or interest(s) outside of work? What is it?
6. What physical malady, or maladies bother you at present, recurrently (if any)?
7. What characteristics of yourself irritate you most?
8. Do you belong to any church in town?
9. What aspect(s) of your physical appearance do you regard as your chief problem, that you dislike most, that you wish you could change or improve?
10. What is the name of the teacher whom you counted as outstanding and most influential to you during your training?
11. Are there any sports you engage in at present, or when there is opportunity?
12. Is there any food or dish that you especially like?
13. With whom in the faculty do you find it most difficult to get along?
14. How much savings do you have at present (within $100)?
15. What is, or was, your father's chief occupation?

2. The amount of personal information that a person knows about each of a group of colleagues covaries with his cathexis for the others.

3. Disclosure-output *to* each colleague covaries with disclosure-intake *from* each colleague.

4. An *S*'s knowledge about each colleague covaries with the amount that each colleague knows about the given *S*.

METHOD

The eight members of a newly organized college of nursing, together with the Dean, served as *S*s in the study. Mean age of the *S*s was 36, with a range from 27 to 45. Several of the *S*s had known one another before joining the present faculty, while others were strangers prior to their appointments. They had worked together for about 10 months at the time the study was undertaken.

Self-disclosure-output was measured in the following manner. Each *S* was interviewed by the investigator, in order to obtain the information that was called for in the questionnaire shown as Table 1. Since the questions do not presume extreme intimacy, the *S*s can be assumed to have been frank in disclosing themselves to the interviewer. Each *S* was then asked to indicate to which of her colleagues she had disclosed information about each item. It was thus possible to determine the total amount of disclosure output for each *S*, and the varying amounts that each *S* had disclosed to each colleague. *Disclosure-intake* was computed from the same data, as the corresponding sums of information directed at a given *S* by her colleagues.

Knowledge was determined by asking each *S* to tell the investigator what she knew, from the questionnaire, about each of her colleagues. A score of 1 was given to each bit of knowledge claimed by an *S* about a colleague which agreed with what the colleague had told the investigator about herself. Similarly, the amount that others knew about a given *S* was obtained from the same data.

Cathexis for the other person was determined by paired comparisons in terms of the criterion, "liked best as a close friend." This procedure produced a rank-order of preference for each *S*, with the best-liked colleague receiving the largest number of choices, and the least-liked colleague receiving the lowest number.

TABLE 2
RELATIONSHIPS AMONG INTERPERSONAL CATHEXIS, DISCLOSURE, AND KNOWLEDGE

Subjects (Ranked in descending order of average cathexis by others)	(1) Total No. of Items Disclosed by *S* to Colleagues	(2) Total No. of Items Disclosed by Colleagues to *S*	(3) Rho, No. of Items Disclosed by *S* to Colleagues vs. *S* Liking of Colleagues	(4) Rho, Disclosure Output to Colleagues vs. Intake from Colleagues	(5) Total No. of Items Known by *S* About Colleagues	(6) Total No. of Items Known by Colleagues about *S*	(7) Rho, No. of Items Known by *S* about Colleagues vs. *S*'s Liking of Colleagues	(8) Rho, No. of Items Known by *S*, about Colleagues vs. Items Known by Colleagues About *S*
Mary	38	65	.79*	.81*	36	39	.71*	.80*
Ann	51	42	.88**	.57	27	29	.23	.34
Ruth	66	57	.72*	.65*	43	37	.49	.86**
Betty	55	58	.76*	.83**	41	27	.67*	.68*
Jean	52	40	.99**	.89**	23	26	.81*	.77*
Rose	36	34	.65*	−.14	23	22	.82*	.04
Suzy	44	46	.79*	.79*	30	44	.52	.65*
Dawn	69	51	−.14	.57	36	41	.11	.83**
Winny	23	41	.05	.80*	33	27	.34	.87**

Note.—Rho = .64 for the .05 level (*) and .83 for the .01 level (**) with $N = 8$.

RESULTS

Significant rank order correlations ($P \leq .05$) between the amount disclosed to each colleague and the degree of liking for the colleague were found for seven of the nine *Ss*. (see Table 2, Column 3). The first hypothesis is thus confirmed.

Column 7 in Table 2 presents the data relating knowledge of others and other-cathexis. Rho's significant at or beyond the .05 level were found in four cases between amount known and liking. The data thus lend support to the second hypothesis.

The relationships just established between self-disclosure, cathexis, and knowledge about others suggest that the nine *Ss* tended to develop pairs of mutual closeness and intimacy. This "dyadic effect" is more clearly demonstrated by further data in Table 2. Column 4 shows that in six cases, significant rho's ($P \leq .05$) were found between disclosure-output to colleagues, and disclosure-intake from colleagues. Correlation between the amount each *S* knew about her colleagues and the amount that her colleagues knew about her is shown by the seven significant rho's reported in Column 8, Table 2. The *Ss* tended to get disclosure back in proportion to what they gave—a lot for a lot, and a little for a little. Similar considerations applied for the relationship between "knowing" and "being known."

DISCUSSION

The present data show that liking, self-disclosure, disclosure-intake, knowing, and being known are interrelated. They show further that despite the existence of a formal role-structure, viz.: a dean, various levels of academic rank, different departments, etc. dyads had tended to develop with varying but largely mutual degrees of intimacy. The data do not show, however, whether liking precedes disclosure, disclosure precedes liking, or whether both of these factors are determined by something else altogether.

Provisionally, it may be suggested that once contact has been made between two persons, they proceed to "uncover" themselves one to the other at a mutually regulated pace. If it is generally true that intimate self-disclosure begets intimate self-disclosure, while impersonality begets impersonality, then certain implications follow for a number of areas of interpersonal endeavor. For example, it might prove to be true that therapists, in order to maximize disclosure in their patients, will be obliged to go beyond impersonal "technique" and "be themselves," i.e., disclose what *they* are experiencing during the therapy hour as freely to their patients as they expect the latter to disclose to them. Rogers (1958, p. 12) already has intimated something of the sort. Possibly similar considerations apply in such fields as teaching, nursing, etc.

Two *Ss* who did not show a significant rho between cathexis for others and disclosure warrant special attention. One of these *Ss* (Dawn) was the highest discloser, while the other (Winny) was the lowest discloser in the group. Both fell at the "least-liked" end of the average cathexis ranks. Direct observation showed that both *Ss* behaved consistently in most situations—the high discloser revealing much about herself without regard to social context, or the interest of the listeners, and the low discloser withholding her personal life even from the scrutiny of those colleagues whom she liked. If being liked by others may be viewed as a rough index of interpersonal competence, (cf. Jourard, 1958, pp. 164–165) then perhaps failure of these *Ss* to vary self-disclosure with cathexis for others betokens contrasting forms of interpersonal (and personal) maladjustment.

It may be objected that the self-disclosure questionnaire items are so trivial that whether or not such information is revealed signifies little about an interpersonal relationship. It can be argued, however, that if a relationship between two people is close, affectionate, and personal, most of the items included in our brief questionnaire will have come up for discussion at one time or another. To paraphrase Schutz (1958, p. 1), "People are interested in one another."

An unsolved question thus far relates to the meaning of self-disclosure as behavior. Why do people vary the amount of self-disclosure in which they engage? If self-disclosure is regarded as a special case of operant behavior (*cf.* Skinner, 1957, pp 185–226), then some answers suggest themselves. Self-disclosure produces consequences, influencing the behavior of others toward oneself for weal or woe. Possibly, then, persons disclose or fail to disclose themselves in accordance with the consequences that they expect to follow. Frequently, the expectancies may not be warranted; in such cases, one would likely find both inaccurate "other-concepts" (cf. Jourard, 1958, pp. 182–191) and either more or less self-disclosure than is appropriate to the situation.

SUMMARY

A self-disclosure questionnaire with 15 items was administered individually to the dean and 8 faculty members of a college of nursing. *Ss* were asked to disclose the answers to these questions about themselves to the investigator, and then to indicate to which of their colleagues they had disclosed each item. The *Ss* were then asked which items of information they knew, from having been told, about each of their colleagues. Finally, the

Ss ordered their colleagues in terms of liking by means of paired comparisons.

It was found that *Ss* tended to vary the amount of disclosure-output to colleagues with degree of liking for colleagues, and to know more about the colleagues whom they liked best than about those whom they liked less well. Finally, evidence was found for structured dyadic relationships such that if an *S* had disclosed much and knew much about a colleague, the other knew much about, and had disclosed much to her.

REFERENCES

JOURARD, S. M. *Personal adjustment. An approach through the study of healthy personality.* New York: Macmillan, 1958.

JOURARD, S. M. & LASAKOW, P. Some factors in self-disclosure. *J. abnorm. soc. Psychol.*, 1958, **56**, 91–98.

ROGERS, C. The characteristics of a helping relationship. *Personnel guid. J.*, 1958, 6–16.

SCHUTZ, W. *FIRO. A three-dimensional theory of interpersonal behavior.* New York: Rinehart, 1958.

SKINNER, B. F. *Verbal behavior.* New York: Appleton-Century-Crofts, 1957.

(Received July 14, 1958)

RELATIONSHIP BETWEEN REAL SIMILARITY AND ASSUMED SIMILARITY WITH FAVORABILITY CONTROLLED[1]

DAVID A. RODGERS

University of California

THE present study tests the hypothesis that subjects who share a number of characteristics in common with a stimulus person tend to assume greater similarity to him than do subjects who have fewer characteristics in common with him. Special attention is given to controlling the influence of favorability on the measurements of assumed similarity.

Reviews of research on social perception (Bruner & Tagiuri, 1954; Taft, 1957) relate accuracy of perception to the dimension of real similarity between judge and judged. In analyzing components of accuracy of perception scores, several writers (Bender & Hastorf, 1953; Cronbach, 1955; Gage & Cronbach, 1955) have pointed out the crucial role of assumed similarity, the tendency for a judge to describe himself and the stimulus object in the same way. Examination of the relationship between real similarity (RS) and assumed similarity (AS) should therefore provide additional information about the nature of social perception.

Studies relating RS and AS or accuracy of perception have tended to ignore the possibility that obtained relationships may be largely determined by the favorability variable (cf. Cronbach, 1955; Gage & Cronbach, 1955). Edwards (1957) has found high positive correlations between the social desirability of personality inventory items and the likelihood that subjects will rate the items as descriptive of themselves. From this relationship, it would be expected that a person's AS score would be higher for stimulus objects he regards favorably than for those he regards less favorably. Such a favorability component in the AS score becomes especially critical when AS or accuracy scores are related to RS measures. If RS consists of such dimensions as ethnic background or social status, the RS score may be highly correlated with favorability of attitude and might be expected to correlate with AS solely because of the favorability component. In the present study, an attempt is made to use RS variables that have relatively equal favorability connotations for all subjects. The favorability component in the AS scores is also parceled out mathematically.

[1] The author is indebted to Audrey Bohm and Judy Chang for assistance in analysis of the data.

PROCEDURE

The *Ss* consisted of 193 college students, 73 male and 120 female. One group of 22 male and 43 female *Ss* was asked to describe a stimulus person (O_1) from information about the following traits (T): sex, age, marital status, religious affiliation, political affiliation, hair color, and eye color. Another group, of 51 male and 77 female *Ss*, was asked to describe a second stimulus person (O_2) from information about the following T: sex, age, marital status, religious affiliation, political affiliation, hair color, eye color, place of birth, and grade average. They made their descriptions using the Gough Adjective Check List (ACL) (1955) in a true–false format. They then indicated their own sex, age, marital status, etc., and described themselves with the ACL.

Both O_1 and O_2 were assigned traits that would have made them typical members of the *S* populations (Table 2). It was hoped that this similarity to the *Ss* and their close associates would cause the Os to be regarded favorably by all *Ss*, so that

TABLE 1
DISTRIBUTIONS OF INDICATED SCORES

Score	Maximum Possible	O_1	O_2	O_3
AS_1	300	197	230	256
AS_2	300	218	242	257
ES_1	300	188	224	247
ES_2	300	207	230	246
FS_1	150	113	125	138
FS_2	150	117	126	133
FO_1	150	116	129	140
FO_2	150	128	135	139

RS and favorability would not be highly correlated. The T variables were sufficiently varied to allow for considerable range in degree and kind of RS. Further variation was assured by assigning different traits to O_2 from those assigned to O_1 (Table 2).

For each T variable, *S*s were divided into high real similarity (RS_H) and low real similarity (RS_L) groups. For example, on the dimension of sex, all *S*s of the same sex as O made up the RS_H group and all *S*s of the opposite sex made up the RS_L group. RS_H and RS_L groups were also determined for three composites of the T variables: total real similarity (RS_t), combining all T variables; salient real similarity (RS_s), combining T variables most likely to produce identification with O (religious affiliation, political affiliation, marital status, grade average); and incidental real similarity (RS_i), combining the remaining variables.

An uncorrected AS score was computed for each *S* by summing the number of adjectives scored the same way on his self-description and on his description of O (Table 1). *S*s were divided at the median of the AS scores into high and low AS groups. An initial test of the hypothesis was made by comparing the AS groupings with the various RS groupings, using chi square tests of significance. The "expected" value in each cell of the contingency tables exceeded 10 in all but two cases.

Gough (1955) lists the 75 ACL adjectives that were judged to be most favorable and the 75 that were judged to be most unfavorable. Using this list, favorability scores (F_S and F_O for *S*s and Os respectively) were determined for each ACL description by summing the number of favorable adjectives scored true and the number of unfavorable adjectives scored false (Table 1). The *F* scores were used to determine the influence that the RS variables had on favorability of attitude and to provide a refined measure of relationship between RS and AS, corrected for the favorability component.

If it is assumed that the favorability dimension extends through all 300 adjectives in the ACL in the same manner that it does throngh the 150 adjectives on the favorability scale, then for any given values of F_S and F_O there will be only a limited range of values that AS can assume. For example, if both F_S and F_O were maximum (150), then all of the adjectives on the favorability scale would necessarily be scored in the same direction and AS would indicate perfect agreement. If it is assumed that no relationship exists between the descriptions of *S* and O except for the favorability component, then the most probable or the expected similarity score (ES) can be determined by the formula

$$ES = 2\left[\frac{F_S F_O}{75} - F_S - F_O + 150\right]$$

ES can then be used to determine whether or not other relationships do exist between the *S* and O descriptions. It corresponds to what Cronbach (1955) has called an assumed-similarity-in-elevation component of AS scores.

Since the sampling distribution of AS changes markedly for different values of F_S and F_O, the following procedure was used to obtain measures of relationship between AS and RS that were relatively uninfluenced by *F*. *S*s were grouped by those having identical (to the nearest whole number) ES scores. Within these groups, they were ranked according to their AS scores and according to their RS_t, RS_s, and RS_i scores (number of traits in common with O). The number of such groups for which the relationship between AS and each of the RS scores was in the predicted direc-

TABLE 2
STIMULUS OBJECT TRAITS AND SIMILARITY TO SUBJECT POPULATIONS

Dimension	Trait		% *S*s with Same Trait	
	O_1	O_2	S_1	S_2
Sex	Male	Female	34	55
Age	35	19	46[a]	23
Marital status	Married	Single	65	—[b]
Religious affiliation	Protestant	Catholic	52	17
Eisenhower or Stevenson supporter	Either[c]	Eisenhower	40	63
Hair color	Brown	Blonde	57	12
Eye color	Blue	Brown	42	48
Birthplace	—	California	—	59
Grade average	—	2.2	—	41[d]

[a] 27 to 43 counted as similar in age.

[b] Information not obtained. Nearly all *S*s in group were unmarried.

[c] Eisenhower in half of descriptions and Stevenson in half. O_1 data were collected just prior to 1956 election; O_2, three months after election.

[d] 2.0 to 2.5 counted as similar grade average.

tion was determined and tested by chi square to see if it exceeded chance expectations. For example, two *Ss* describing O_1 received ES scores of 260. They ranked 1 and 2 on AS and 2 and 1 on RS_t, so were counted as one group with results in the direction opposite from that predicted for the AS-RS_t relationship. Similar comparisons were made between AS and RS_s and RS_i. By the same procedure, the relationships between AS and the individual T characteristics of sex, political affiliation, and religious affiliation were also determined. These comparisons concern essentially Cronbach's assumed-dispersion-on-specific-items (ADI) component of the AS scores. Of *Ss* describing O_1, 13 groups with identical ES scores were obtained. Of *Ss* describing O_2, 37 such groups were obtained.

RESULTS AND CONCLUSIONS

F_O was found to be not significantly related to RS on any single or composite dimension, indicating that the favorability component of RS was adequately controlled. Highly significant relationships were found between F_S and F_O. The product-moment correlation between F_{S_1} and F_{O_1} was .49 ± .08 and between F_{S_2} and F_{O_2} was .47 ± .06, both correlations being significant at beyond the .001 level. Means of the *F* scores were high and the scores were markedly negatively skewed (Table 1), making it inevitable that the correlations between the *F* scores and AS would be positive and high. They ranged from .69 to .76. Corrected for ES, however, the partial correlations drop to near zero (between −.10 and +.10), indicating that no appreciable relationship exists between the *F* scores and the ADI component of AS. Correlations between ES and AS were over .80, although it should be noted that ES and AS are not mathematically independent and would therefore be expected to be positively correlated.

Almost no relationship was found between AS and RS. In 23 comparisons made between AS and the various RS groupings, only one relationship reached statistical significance. AS related significantly in the expected direction at the .02 level to religious affiliation of O_1. One such significance level could be expected by chance and must therefore be largely discounted, especially since no similar relationship was found to religious affiliation of O_2. The refined comparisons between AS and the RS scores failed to produce any significant relationships or promising trends.

The favorability data illustrate the importance of Gage and Cronbach's (1955) emphasis that response sets can markedly influence AS scores. They are consistent with Edwards's (1957) findings that the social desirability variable is a highly significant component of personality assessments made with currently used inventories. The data almost completely fail to support the hypothesis that AS varies with RS when the favorability component is controlled. Other RS dimensions might, of course, produce other results. Within the limitations of these dimensions and of the particular ACL used, however, the data suggest that, when the stimulus object is relatively socially acceptable and has characteristics with which all *Ss* are relatively familiar, AS is determined primarily by the favorability response set of *S* and is almost uninfluenced by RS.

REFERENCES

BENDER, I. E. & HASTORF, A. H. On measuring generalized empathic ability (social sensitivity). *J. abnorm. soc. Psychol.*, 1953, **48**, 503–506.

BRUNER, J. S. & TAGIURI, R. The perception of people. In G. Lindzey (Ed.), *Handbook of social psychology*. Vol. II. Cambridge, Mass.: Addison-Wesley, 1954. Pp. 634–654.

CRONBACH, L. J. Processes affecting scores on "understanding of others" and "assumed similarity." *Psychol. Bull.*, 1955, **52**, 177–193.

EDWARDS, A. L. *The social desirability variable in personality assessment and research*. New York: Dryden, 1957.

GAGE, N. L., & CRONBACH, L. J. Conceptual and methodological problems in interpersonal perception. *Psychol. Rev.*, 1955, **62**, 411–422.

GOUGH, H. G. *Reference handbook for the Gough Adjective Check List*. Berkeley: Univer. California Inst. Personality Assessment & Research, 1955.

TAFT, R. Some correlates of the ability to make accurate social judgments. Unpublished doctoral dissertation, Univer. of California, 1950.

(Received August 23, 1958)

SOCIAL DESIRABILITY AND THE DESCRIPTION OF OTHERS[1]

ALLEN L. EDWARDS

University of Washington

THE relationship between the social desirability scale value of a personality item and the probability that it will be endorsed in self-description has been the subject of several studies (Edwards: 1953, 1957a; Hanley, 1956; Kenny, 1956). In general, these studies, summarized by Edwards (1957b), show that probability of endorsement of an item is a linear increasing function of the social desirability scale value of the item.

The present study was designed to investigate the relationship between the social desirability scale value of an item and the probability that the item will be endorsed not in self-description but rather in describing others under various sets of instructions.

METHOD

The personality items used in the study were those contained in the 128-item Interpersonal Check List (ICL) (Leary), as described by LaForge and Suczek (1955). Equal-appearing interval social desirability scale values for each of the 128 items in the ICL were available from an earlier study (Edwards, 1957a).

A total of 1067 descriptions of others were obtained from three different groups of *Ss* under three different sets of instructions. One group of *Ss*, the Like group, consisted of 57 males and 55 females registered in an introductory speech course at the University of Washington. *Ss* in this group were told to think of someone of the same age and sex that they liked most and to describe this person with the ICL. Another group, the Dislike group, consisted of 73 males and 42 females also registered in an introductory speech course at the University of Washington. *Ss* in this group were told to think of someone of the same age and sex that they disliked most and to describe this person with the ICL.

A third group of *Ss*, the Other group, consisted of 95 males and 73 females enrolled in an introductory psychology course at the University of Washington. *Ss* in this group were told that the objective of their task would be to obtain estimates of the percentage of students who could be accurately described by each of a number of adjectives. They were asked to think of the names of five individuals of the same sex and of approximately the same age that they knew intimately. They were also told that these five individuals should not necessarily be their best friends nor their worst enemies, but simply five individuals whose behavior they had observed frequently and whose personality characteristics they felt they knew quite well. Each *S* was then asked to describe these five individuals with the ICL. A total of 475 male and 365 female descriptions were obtained from this group.

Each of the *Ss* in the Other group had taken the Social Desirability Scale (SDS) approximately one to two weeks prior to making their descriptions of others. The SDS has been described by Edwards (1957b) and consists of 39 items drawn from the Minnesota Multiphasic Personality Inventory (MMPI) keyed for socially desirable responses. A socially desirable response is defined as either a True response to an item with a socially desirable scale value or a False response to an item with a socially undesirable scale value. The higher the score on the SDS, the more frequently an *S* has given socially desirable responses in self-description. *Ss* in the Other group were subdivided into two groups, a High SD (HSD) group and a Low SD (LSD) group. The division of the *Ss* into these two groups was made by splitting the distribution of SD scores at approximately the median. This division resulted in a group of 46 HSD and a group of 49 LSD males. For the females, the HSD group contained 36 members and the LSD group 37.

The data collected were analyzed separately for each sex group under each set of instructions. For each of the 128 items in the ICL the proportion checking or endorsing each item in each group was found. These proportions are considered as estimates of the probability that a given item will be attributed to another person under the particular instructions given to the group of *Ss*. Product-moment correlations were then obtained between probability of endorsement and social desirability scale value over the 128 items separately for each sex group and under each condition of instruction. All tests of significance, reported later, with respect to these correlation coefficients were made using Fisher's z transformation for r.

RESULTS AND DISCUSSION

Table 1 summarizes the results of the study. All correlations except those between self-descriptions and social desirability scale values were obtained

[1] This study was supported by a grant from the graduate school of the University of Washington.

TABLE 1

MEANS AND VARIANCES OF ITEM PROBABILITIES FOR THE INTERPERSONAL CHECK LIST AND CORRELATIONS BETWEEN ITEM PROBABILITIES AND SOCIAL DESIRABILITY SCALE VALUES OF THE ITEMS OBTAINED UNDER DIFFERENT INSTRUCTIONS

Group	Males				Females			
	n	$\bar{X}$	s^2	r	n	$\bar{X}$	s^2	r
Self	66	.48	.087	.82	67	.46	.090	.84
Dislike	73	.43	.071	−.38	42	.45	.065	−.13
Like	57	.46	.104	.93	55	.42	.123	.95
Other	95	.41	.064	.84	73	.46	.065	.86
HSD	46	.42	.079	.87	36	.42	.091	.90
LSD	49	.49	.051	.79	37	.43	.064	.85

in the present study. The correlations between probability of endorsement of an item in self-description and social desirability scale value of the item are based upon data obtained in an earlier study (Edwards, 1957a). These correlations for males and females, .82 and .84, respectively, are quite similar in value to the corresponding correlations of .84 and .86 obtained in the present study when males and females are asked to describe others.

When *S*s are asked to describe someone they like, correlations of .93 for males and .95 for females are obtained between probability of item endorsement and social desirability scale value. Both of these correlations are significantly higher ($P < .0003$) than the corresponding correlations between probability of endorsement in self-description and social desirability scale value.

When *S*s are asked to describe someone they dislike most, the correlations between probability of endorsement and social desirability scale value are −.38 for males and −.13 for females. The male value represents a highly significant departure from zero in a negative direction ($P < .0001$). For the females, the probability of a negative correlation equal to or less than −.13 is approximately .07. These correlations, although in the expected direction, are not of the magnitude of those obtained under other sets of instructions.

Knowing the social desirability scale value of an item, one can predict quite well the probability that the characteristic will be attributed to someone liked, whereas the corresponding prediction for a disliked person would be quite poor and subject to considerable error. This finding suggests that there is less agreement as to the characteristics associated with a disliked person while there is greater agreement as to the characteristics associated with a liked person. In other words, one's dislike of another person may be associated with the presence or absence of rather specific traits relatively unrelated to the normative social desirability scale value of these traits. Evidence on this point is available from the differences in the variability of the probabilities of endorsement for liked and disliked persons. With greater agreement in the characteristics associated with liked persons than with disliked persons, the variance of the probabilities for liked persons should be greater than the variance for disliked persons. The F test for homogeneity of these variances for males is .104/.071 = 1.46 and for females is .123/.065 = 1.89. These two values of F, each with 127 and 127 degrees of freedom, are both statistically significant ($P < .05$).

*S*s in the Other group were subdivided into HSD and LSD groups because it was expected that differences would be found in the manner which these two groups describe others. Research with the SDS, reported by Edwards (1957b), for example, has shown that HSD *S*s are those who tend to give primarily socially desirable responses in self-description on a variety of personality inventories, whereas LSD *S*s are less influenced by the tendency to give socially desirable responses. If the same tendency is operating when HSD *S*s describe others, then for items with high social desirability scale values, the HSD probability should be greater than the LSD probability, that is, a larger proportion of the HSD group should attribute these items to others than in the case of the LSD group. On the other hand, for items with low social desirability scale values, the proportion of LSD *S*s attributing these items to others should be greater than the proportion in the HSD group. As a result, the over-all variability in the item probabilities for the HSD group should be greater than the corresponding variability for the LSD group. With a restriction in range of item probabilities, the correlation between the item probabilities and social desirability scale values should also be somewhat lower for the LSD group than for the HSD group.

The data are consistent with the above analysis. For the male *S*s, for example, there were 8 items for which the HSD and LSD probabilities were equal. Of the remaining 120 items, 54 have social desirability scale values of 5.0 or greater, and on 52 of these 54 items the HSD probability is greater than the LSD probability. On the 66 items with social desirability scale values less than 5.0, the LSD probability is greater than the HSD probability 54 times out of the possible 66 times that it could be. Similar results are found for the female *S*s. The female HSD and LSD groups had equal probabilities on 7 items. On the remaining 52 items with social desirability scale values of 5.0 or greater, the HSD probability is greater than the LSD probability 47 times out of 52. The

reverse trend is true for the 69 items with social desirable scale values less than 5.0. On 60 of these 69 items, the LSD probability is greater than the HSD probability. For both males and females, then, there is a definite tendency for HSD *Ss* to be more likely to attribute socially desirable characteristics to others than LSD *Ss*. On the other hand, LSD *Ss* are more likely to attribute socially undesirable characteristics to others than HSD *Ss*.

A test of significance with respect to the differences in the variances of the item probabilities for the HSD and LSD males is given by $F = .079/.051 = 1.55$. The corresponding test for the HSD and LSD females gives $F = .091/.064 = 1.42$. These two values of F each with 127 and 127 degrees of freedom are both significant ($P < .05$) and it can be concluded that both the HSD male and female groups show greater variability in their proportions of item endorsements than the corresponding LSD groups.

Tests with respect to the differences in the correlations between probability of endorsement and social desirability scale value for HSD and LSD groups also give significant results. The correlation of .87 for HSD males is significantly greater ($P < .05$) than the value of .79 for LSD males, and the correlation of .90 for HSD females is also significantly greater ($P < .05$) than the corresponding value of .85 for LSD females.

SUMMARY

This study reports upon the correlations found between the social desirability scale values of the 128 items in the ICL and the corresponding probabilities of item endorsements when the ICL is used to describe oneself or others. High and positive correlations were found between probability of item endorsement and social desirability scale value when *Ss* were asked to describe themselves and when they were asked to describe someone liked. The correlation obtained between social desirability scale value and probability of item endorsement when *Ss* described someone disliked was negative but of low magnitude. It was suggested that one's dislike of another may be associated with the presence or absence of rather specific traits relatively unrelated to the normative social desirability scale values of these traits.

Descriptions of others were also obtained from an HSD group, characterized as being relatively high in their tendencies to describe themselves in socially desirable terms, and from a LSD group, characterized as being relatively low in this respect. For items with high social desirability scale values, it was found that the probabilities obtained from the HSD group were consistently greater than those obtained from the LSD group. On the other hand, for items with low social desirability scale values, the probabilities obtained from the LSD group were consistently greater than those obtained from the HSD group.

REFERENCES

EDWARDS, A. L. The relationship between the judged desirability of a trait and the probability that the trait will be endorsed. *J. appl. Psychol.*, 1953, **37**, 90–93.

EDWARDS, A. L. Social desirability and probability of endorsement of items in the Interpersonal Check List. *J. abnorm. soc. Psychol.*, 1957, **55**, 394–396. (a)

EDWARDS, A. L. *The social desirability variable in personality assessment and research* New York: Dryden, 1957. (b)

HANLEY, C. Social desirability and responses to items from three MMPI scales: *D*, *Sc*, and *K*. *J. appl. Psychol.*, 1956, **40**, 324–328.

KENNY, D. T. The influence of social desirability on discrepancy measures between real self and ideal self. *J. consult. Psychol.*, 1956, **20**, 315–318.

LAFORGE, R., & SUCZEK, R. F. The interpersonal dimension of personality: III. An interpersonal check list. *J. Pers.*, 1955, **24**, 94–112.

LEARY, T. *The Interpersonal Check List.* Berkeley, Calif.: Psychological Consulting Service. (Undated)

(Received August 23, 1958)

VALUE AND THE ACCENTUATION OF JUDGED DIFFERENCES: A CONFIRMATION

H. TAJFEL AND S. D. CAWASJEE

University of Oxford

As was pointed out in a previous paper (Tajfel, 1957), the numerous experiments concerned with the phenonmenon of perceptual overestimation due to motivational or value variables fall into two classes. In one group, there is no relationship of any kind between the physical magnitudes of the stimuli judged by the *Ss* and the value of these stimuli to them; in the second group there is a consistent relationship between the physical magnitudes and value. For example, larger coins

tend to be more valuable than smaller ones; but there is no relationship between the size of a swastika and the degree of its emotional relevance to the subject (Klein, Schlesinger, & Meister, 1951). A review of the evidence (Tajfel, 1957) showed that most of the positive results were obtained in the "relevant" group of experiments; most of experiments in the "irrelevant" category yielded negative results. The explanation of this was sought in the assumption that when physical magnitudes and value vary concomitantly in a series of stimuli, the differences between the stimuli of the series are judged larger than in a physically equivalent series where there is no consistent relationship between value and the physical dimension under judgment. Overestimation was conceived as a by-product of this stretching of the response scale which would result in the stimuli at the large and small ends of the series being judged as larger and smaller respectively than the physically equivalent stimuli in a control "neutral" series.

McCurdy (1956) came to a rather similar conclusion in a paper concerned with the bearing of the concept of schemata on "coin perception studies," and Bruner and Rodrigues (1953) concluded from an experiment conducted some years ago that the concept of "relative" overestimation was much more helpful to clarify the results obtained than a simple notion of an "absolute" overestimation of valued stimuli. In some experiments recently performed by one of the writers (Tajfel, in press), it was also found that an association of value with either end of a series of weights led to an extension of the response scale (larger judged differences between the stimuli lying at the opposite extremes of the series), while there was no clear-cut evidence for an overestimation of the stimuli associated with value.

Coins took a rather undue prominence in experiments in this field; since they served again as stimuli in the present experiment, it may be worth adding that the writers do not consider the judgment of size of coins as a problem of intrinsic importance and interest. The results obtained have, however, some interesting implications for phenomena of much greater generality: namely, the effects of abstract attributes of stimuli on the judgment of their physical characteristics. As has been argued elsewhere (Tajfel, 1959), value is no more than one instance of an abstract attribute, and coins are only one instance of its operation.

In "natural" conditions the denomination of a coin is judged through a swift and implicit comparison with other members of the series. The identifying decision may be based on differences between the coins in color, design, or simply on the fact that the value is specified. When none of these cues exist, and size is the main differential, a weighing in terms of size cues must become the prominent feature of the process of judgment. Again, in many (possibly most) "natural" situations one is not allowed to base one's decision on a comparison between two members of the series present simultaneously; a coin must be quickly and correctly identified when it is seen in the absence of others. The concern is not with the size of the coin expressed in terms of absolute units of measurement; it is with the difference between this coin and those near to it in the order of size.

In the paper referred to above (Tajfel, 1957), theoretical reasons were given why this situation should lead to an accentuation of judged differences. This "implicit" accentuation should manifest itself explicitly when all the elements of the usual situation are maintained, with the addition of a request to the *S*s to verbalize in quantitative terms their estimates of the size of coins. The *S*s, when confronted with this request should, on the present assumptions, judge the differences between the coins to be significantly larger than the corresponding differences between the control stimuli. An assumption of simple overestimation of valuable stimuli (especially with the corollary that the amount of overestimation is a function of the amount of value) would not contradict this statement directly; but it would lead to the prediction that all the valuable stimuli should be overestimated, the accentuation of judged differences between them being the by-product of different amounts of overestimation. This may be viewed as a rubber band placed on a ruler, transported some distance to the right, and then possibly stretched outwards from its middle in both directions. The alternative predicted here is of the rubber band remaining in its original position and similarly stretched.

METHOD

The British coins of 2 shillings and 2 shillings and sixpence (2/- and 2/6d) offered the advantage of differing, for all practical purposes, in size only. The tails design consists of coats of arms. These, though not identical in both coins, are very similar; in addition, coins of the same denomination minted at various dates differ in the design of their coats of arms. The value of the coins is not stated in numbers; the 2/- coin displays the words "one florin," and the 2/6d coin the words "half crown" in the middle of an identical Latin inscription in small letters running round the circumference in both coins. Heads in both coins are identical: they display the profile of the monarch reigning at the time of minting.

Selection of control stimuli presented some difficulty, as they had to achieve a maximum similarity with the coins, without being at the same time con-

TABLE 1
MEAN ESTIMATES (IN INCHES) OF LENGTH OF DIAMETER OF STIMULI

	Current Coins					Control Coins				
	Heads		Tails		Mean	Heads		Tails		Mean
	A	B	A	B		A	B	A	B	
Large stimuli	1.42	1.45	1.53	1.57	1.49	1.36	1.47	1.26	1.30	1.35
Small stimuli	1.15	1.18	1.21	1.26	1.20	1.16	1.23	1.08	1.12	1.15
Diff.	0.27	0.27	0.32	0.31	0.29	0.20	0.24	0.18	0.18	0.20

Note.—A—large stimuli presented first; B—large stimuli presented second.

TABLE 2
MEAN PERCENTAGE DIFFERENCES BETWEEN ESTIMATES OF LARGE AND SMALL STIMULI

	Current coins		Control Coins	
	Heads	Tails	Heads	Tails
Large stimulus first	23.5	26.4	17.6	17.5
Large stimulus second	22.9	23.5	18.7	15.8
Over-all means	24.2		17.4	

TABLE 3
ANALYSIS OF VARIANCE OF PERCENTAGE DIFFERENCES BETWEEN THE ESTIMATES OF LARGE AND SMALL STIMULI

Source	*SS*	*df*	*MS*	*F*
Value (Current vs. control coins)	2231.12	1	2231.12	19.55*
Sides (Heads vs. Tails)	1.25	1	1.25	—
Order of presentation	53.46	1	53.46	—
V × S	127.68	1	127.68	—
V × O	25.78	1	25.78	—
S × O	79.38	1	79.38	—
V × S × O	.67	1	.67	
Error	21913.22	192	114.13	

* $p < .001$.

fused with them. The control stimulus for the 2/- coin was an 1896 South African coin with the profile of Krueger on one side and a coat of arms with an inscription on the other. The coin, originally very slightly larger than the 2/- piece, was filed down to size. The control stimulus for the 2/6d coin was an 1834 2/6d coin (which still has official currency, but is extremely rare and practically unknown) with the profile of William III on one side and a coat of arms on the other. The texture and brightness of the current and control coins were identical. The profiles on the two current coins were not the same in order to approach, as far as possible, the lack of similarity between Krueger and William III. The diameter of the smaller stimuli was 1.125 in.; of the larger—1.250 in.

Two hundred *S*s each made one judgment of either

TABLE 4
ANALYSIS OF VARIANCE OF SIZE ESTIMATES OF THE LARGER STIMULUS

Source	*SS*	*df*	*MS*	*F*
Value (Current vs. control coins)	102.39	1	102.39	4.57*
Sides (Heads vs. Tails)	.86	1	.86	—
Order of presentation	14.31	1	14.31	—
V × S	79.51	1	79.51	3.55
V × O	2.02	1	2.02	—
S × O	.86	1	.86	—
V × S × O	2.18	1	2.18	—
Error	4301.16	192	22.40	

* $p < .05$.

the experimental or the control pair of stimuli, presented successively. *E* held the coin flat in the palm of his hand at a distance of about 16 in. from the *S*'s eyes and asked him to estimate the diameter in terms of tenths of an inch. Those *S*s who worried about the difficulty of the task were told to stop worrying and try as best they could. The *S*s were recruited for several weeks *au hasard des rencontres:* in restaurants, hotels, canteens, libraries, colleges, coffee (nonalcoholic) parties, trains, etc. No one who had the opportunity of hearing someone else's estimates was used as an *S*. The 200 *S*s were divided into 8 groups of 25 each: half judged the current coins, half the control coins; half judged heads, half tails; half judged the larger stimulus first, and half the larger stimulus second. Assignment of *S*s to groups was randomized.

RESULTS

Results in terms of mean estimates are shown in Table 1. Table 2 shows the judged size relationships between the two coins. The basis of scoring here was for each *S* his judgment of the diameter of the smaller coin; the difference between his estimates of the larger and of the smaller coin was then expressed as the percentage of the judged diameter of the smaller coin. The actual difference between the two coins is 11.1% of the diameter of the smaller coin.

As can be seen from Table 3, an analysis of variance conducted on these scores shows a highly

significant difference in the expected direction between the judged size relationships of the two current coins and the two control stimuli respectively. None of the other variables approach significance.

Analysis of variance was then conducted for the actual estimates of the larger and smaller stimuli separately. None of the variables reaches significance for the smaller stimulus. There is a significant overestimation of the larger coin in relation to the larger control stimulus (see Table 4).

DISCUSSION

The hypothesis of an accentuation of judged differences between the stimuli in a series varying concomitantly in size and value finds very clear support in the data. The possibility that the small differences in design between the current coins and the control stimuli may have somehow contributed to the results seems extremely remote in view of the fact that they are equally clear-cut for both sides of the stimuli.

The evidence for overestimation per se as function of value is more ambiguous; it is moderately consistent for the larger coin, and nonexistent for the smaller one. The expectation was that some amount of overestimation would be found for the larger coin, and some amount of underestimation for the smaller one. The latter prediction was not borne out. It is possible, however, that the smaller coin was not underestimated because it is not the smallest member of the series to which it belongs. This series consists of four coins (sixpence, one shilling, two shillings, and two shillings and sixpence) which differ mainly in size and value. The existing evidence that all stimuli of a series, present or absent, have an effect on the judgment of each individual member of the series, and that this effect is a function of the extent to which the stimuli are perceived as "belonging" together, lends support to this view.

SUMMARY

Subjects were requested to estimate the diameter of two coins differing in size and value, and presented successively. Unfamiliar coins of identical sizes, similar design, and identical texture and brightness were used as control stimuli. The results show that judged differences between the current coins are significantly larger than the judged differences between the control stimuli, and provide evidence that this phenomenon is much more consistent and clear-cut than simple overestimation due to value.

REFERENCES

BRUNER, J. S., & RODRIGUES, J. S. Some determinants of apparent size. *J. abnorm. soc. Psychol.*, 1953, **48**, 17–24.

KLEIN, G. S., SCHLESINGER, H. J., & MEISTER, D. E. The effect of values on perception: An experimental critique. *Psychol. Rev.*, 1951, **58**, 96–112.

MCCURDY, H. G. Coin perception studies and the concept of schemata. *Psychol. Rev.*, 1956, **63**, 160–168.

TAJFEL, H. Value and the perceptual judgment of magnitude. *Psychol. Rev.*, 1957, **64**, 192–204.

TAJFEL, H. Quantitative judgment in social perception. *Brit. J. Psychol.*, 1959, **50**, 16–29.

TAJFEL, H. The anchoring effects of value in a scale of judgments. *Brit. J. Psychol.*, in press.

(Received August 23, 1958)

THE INFLUENCE OF AN AFFECTIVE SET ON DISYLLABLE RECOGNITION THRESHOLDS[1]

ROBERT O. HATFIELD[2]

University of Denver

AMONG the "new look" perceptionists there is still considerable controversy over the relative merits of word frequency, set, and motivational factors as alternative ways of explaining perceptual behavior under conditions of subliminal sensitivity (Cowen & Obrist, 1958; Howes & Solomon, 1951; Spence, 1957). The present experiment was designed to test the hypothesis that there is a relationship between motivational factors and perception. Electric shock was employed to induce anxiety provoking conditions in relation to the perception of certain verbal symbols, and changes in their speed of perception were noted.

An incidental sample of 86 college students, 43 male and 43 female, were tested individually.

In order to avoid the difficulties involved in

[1] This paper, which was read at the APA Convention in Washington, D. C., August 1958, is a report adapted from a dissertation submitted in partial fulfillment of the requirements for the Ph.D. degree at the University of Denver.

[2] Now with the Child Guidance Clinic of Southeastern Connecticut, New London.

TABLE 1
LIST OF STIMULUS DISYLLABLES

Disyllable	Mean Association Value[a]	Letter Legibility Rank[b]
Gokem	1.27	57
Latuk	1.26	66
Polef	1.30	59
Tarop	1.24	45
Xylem	1.24	48
Zumap	1.28	41

[a] Indicates mean frequency of continuous associations in a 60-sec. interval.

[b] Indicates rank legibility according to a weighting of individual letters making up each syllable.

word frequency, and to deal effectively with selective reporting sets upon the part of subjects, relatively meaningless disyllables were used. Even nonsense syllables vary in their formal structural characteristics, however, so dissyllables were selected which had been standardized as to association value in a previous experiment (Noble, 1952). Configurational cues and differences in legibility of different letters in the alphabet were controlled by typing all the letters in capitals and choosing disyllables of approximately equal letter legibility (Tinker, 1928). For a list of disyllables used see Table 1. In addition, in order to eliminate the possibility of errors from the influence of stimulus generalization, practice and fatigue effect, list order preferences, and any differences in the structural characteristics of the syllables themselves, all syllables were presented in a randomized order, and all syllables were given an equal chance of being chosen as shock syllables for any given subject.

The disyllables were given need-avoidant or taboo qualities through coupling them with electric shock. Electric stimulation was set at a point half way between minimum sensitivity or the point where *S* could just sense it and that at which it was rated as intolerably painful, and which also aroused an above average GSR. In every case the electric stimulation was increased gradually during the course of the experiment. As *S* became adapted to the intensity of the current it was raised enough to insure the maximum unconditioned response possible from the *S*. In most cases the final amount was greater than that rated as intolerable by *S* at the beginning of the experiment. The range of intensities for all *Ss* for the experiment as a whole was from 1000 to 5000 microamperes.

The apparatus consisted of a tachistoscope, electric stimulator, and galvanometer plus accessory equipment.

The experiment was divided into four stages or periods: pretest, acquisition, transition, and final test. In the pretest period the range of visual acuity of each *S* was determined in terms of recognition thresholds with the tachistoscope, and some practice was given in the nature of the task involved.

The acquisition period consisted of six trials of six presentations of each syllable on the screen. In this period *Ss* were conditioned with electric shock to one half the syllables while the other half was presented without shock. The CS–US interval was two seconds and the reinforcement pattern was a partial one of 33%. GSRs were recorded on the nonreinforced trials as a measure of conditioning. During this period *Ss* were instructed not to report aloud the syllables presented in order to avoid the chance that they might associate shock with their verbal report and thus influence their response preference in the final test period. Conditioned responses were defined operationally as those of greater amplitude than the average of the nonshock syllables in the same series of presentations.

In order to speed up the process of conditioning *Ss* were told in advance on which syllables they would be shocked. This procedure tended to produce verbal conditioning (Cook & Harris, 1937) which supplemented the direct conditioning.

In the transition period, of four to six trials, the duration of exposure was gradually shortened until *S*'s threshold was exceeded. This was done to make this period as much like the final test period as possible so that *S* would not lose the expectancy of shock.

In a preliminary study it was discovered that exposure of .01 sec. did not make the task sufficiently difficult for most *Ss*. Therefore, a variac was placed in the circuit to cut down the voltage of the current to the tachistoscope lamp. In addition, the light was further dimmed by adjusting a diaphragm on the tachistoscope to graduated openings on its rim. The diaphragm was graduated in degrees by comparison with a standard foot-candle meter, set in a direct line with the mirror of the tachistoscope one yard

TABLE 2
A COMPARISON OF MEAN RECOGNITION THRESHOLDS FOR THE SHOCK AND NONSHOCK DISYLLABLES IN FOOT-CANDLES OF LIGHT NECESSARY FOR CORRECT RECOGNITION

Treatments[a]	*N*	*M*	*SD*	*CR*	*p*
Shock	65	46.3	17.8	3.8	<.01
Nonshock	65	51.1	21.1		

[a] Refers to the Criterion group.

away from it. A light scale was worked out in degree units for the diaphragm such that for any given voltage supplied by the variac the light transmitted was proportional in foot-candle power units. Previous research has shown that with exposure times shorter than 50 ms. the effective stimulation equals the product of time and light intensity.

In the final test period the light was gradually increased in a series of trials until *S* was able to recognize all the syllables successfully.

Instructions to the *S* included: orientation as to the nature of the experiment, reassurance as to the intensity of the shock, the importance of sitting still during GSR readings, and the procedure in the various test periods.

A treatment by *Ss* design was employed in which each *S* served as his own control. A pair of mean recognition threshold scores for each *S* was obtained using a modified method of limits in the ascending series. A criterion of conditioning was set up for the inclusion of *Ss* in the final experimental group. This required each *S* to give a conditioned response over 50% of the time to the shock syllables on the nonreinforced trials in the transition period.

RESULTS

The difference between the mean thresholds for the shock and nonshock syllables for the group which met the criterion of conditioning was 4.8 foot-candles (Table 2). Significantly lower thresholds were thus found for the shock than for the nonshock syllables. For *Ss* who did not reach the criterion of above chance conditioning this difference was less (3.0 foot-candles, $p = .05$), though in the same direction.

A correlation coefficient between the degree of conditioning and the magnitude of threshold differences (disregarding the sign) showed a low positive but significant correlation ($r = .27$, $p = .05$).

REFERENCES

COOK, W. S., & HARRIS, R. E. The verbal conditioning of the galvanic skin reflex. *J. exp. Psychol.*, 1937, **21**, 202–210.

COWEN, E. L., & OBRIST, P. A. Perceptual reactivity to threat and neutral words under varying experimental conditions. *J. abnorm. soc. Psychol.*, 1958, **56**, 305–310.

HOWES, D. H., & SOLOMON, R. L. Visual duration thresholds as a function of word probability. *J. exp. Psychol.*, 1951, **41**, 401–410.

NOBLE, C. E. An analysis of meaning. *Psychol. Rev.*, 1952, **59**, 421–430.

SPENCE, D. P. A new look at vigilance and defense. *J. abnorm. soc. Psychol.*, 1957, **54**, 103–108.

TINKER, M. A. The relative legibility of letters, digits, and certain mathematical signs. *J. gen. Psychol.*, 1928, **1**, 472–496.

(Received September 29, 1958)

METHOD IN PHONETIC SYMBOLISM EXPERIMENTS[1]

ROGER BROWN
Massachusetts Institute of Technology
AND RONALD NUTTALL
Harvard University

IN 1955 Brown, Black, and Horowitz reported an experiment in which English speaking *Ss* were asked to guess the meanings of 21 pairs of contrasting words presented in three foreign languages—Chinese, Czech, and Hindi. For each language the number of correct choices was significantly greater than the chance expectation. In two experiments on phonetic symbolism reported since 1955 (Brackbill & Little, 1957; Maltzman, Morrisett, & Brooks, 1956) results were obtained which in part agree with the earlier finding and in part do not. However, where the earlier result has not been replicated there have been certain changes of experimental procedure. In order to find out whether these changes of procedure affect the success with which *Ss* can guess the meanings of unfamiliar words the present investigators have used the original list with three groups of *Ss* in three different ways. The results support the hypothesis of phonetic symbolism and suggest a more precise formulation of it.

In the study of Brown, Black, and Horowitz (1955), pairs of contrasting English words (e.g., *dark–light*, *fat–thin*) were ranged down one side of a page before the *S*, and on the other side appeared pairs of foreign language equivalents.

[1] This investigation was supported by a research grant from the Laboratory of Social Relations, Harvard University.

Words within a pair were randomly arranged so that while *S*s knew that a given English word matched one or the other of the two foreign words standing opposite, they could not tell from the arrangement which of the two it was. *S*s heard the words pronounced (the foreign words by native speakers) in the order in which they appeared on the sheet. We shall call this the English-Foreign-Pairs method of presentation.

In the study of Maltzman et al. (1956), words were used from English, Japanese, and Croatian. *S*s were given a stimulus word and two response words from which to choose the equivalent of the stimulus word. When the task included at least one English word *S*s scored at levels significantly above chance, but they failed to score at this level when the task involved Croatian stimulus words and pairs of Japanese response words. We shall refer to presentations involving two foreign languages as the Foreign-Foreign method.

Brackbill and Little (1957) used a list of 50 words constituting a random sample of concepts of high frequency in usage. This was not a list restricted to contrasting pairs but included such terms as *when, first, this, etc.* The English forms were translated into Chinese, Japanese, and Hebrew. *S*s were presented with two words at a time (the two being from different languages) and were asked to judge whether they were the same or different in meaning. The experimenter told *S* that half of the pairs were the same. In making up their pairs the authors arrived at the incorrect matchings by random assignment of the words remaining after the correct pairs had been matched, and so words that were different were not necessarily antonymic or even contrasting. The four languages were combined in all possible sets of two to yield six different collections of paired words. With this procedure *S*s were able to guess with better than chance success for English-Hebrew (53.0%), Chinese-Japanese (54.8%), and Hebrew-Japanese (52.3%), but with success at or even significantly below chance levels with English-Japanese (50.3%), English-Chinese (49.9%), and Chinese-Hebrew (48.1%). Even on the successful sets the percentages are all below those in the study of Brown et al. (1955) where they were 55.7% for English-Czech; 58.9% for English-Chinese; 59.6% for English-Hindi. We will call the method used by Brackbill and Little (1957) the Same–Different procedure.

This word-matching design has involved several other procedural variations. There is, for example, the matter of whether the words are to be presented by an audio-visual method or in an exclusively visual or exclusively auditory way. In the experiment of Maltzman et al. (1956), the words were presented by visual means alone; in the experiment of Brown et al. (1955), by audio-visual means, and for a small comparison group by exclusively visual means. The latter two results were much the same. Brackbill and Little (1957) systematically varied the three techniques and found significant interaction with the particular language pairs, but the totals taken across language pairs for the three methods of presentation did not vary by more than about 1%. In the present experiment the audio-visual presentation has been used for all three groups.

More important than the mode of presentation is the selection of the word list. All three studies are agreed on the essential point that the words must be selected in English by persons having no knowledge of the foreign languages into which they are to be translated. Brackbill and Little (1957) have made telling criticisms of certain unclarities and inconsistencies in the selection criteria used by Brown et al. (1955). Not all of the words on this original list are antonyms (e.g., *gold–iron*) and not all of them name sense dimensions (e.g., *gold–iron*). Nevertheless, the present experiment works with the original list so as to have materials that yield a sizable phonetic symbolism effect when presented in the original manner. With this list it will be possible to determine whether certain changes of procedure reduce or eliminate the evidence of phonetic symbolism. Two minor changes were made in the original lists. Only two of the foreign languages were used: Chinese and Hindi. To save time one language was omitted and Czech was chosen because the results with Czech had been smaller than with Chinese and Hindi and the present experiment is intended to maximize the phonetic symbolism effect insofar as the various procedures permit. Finally, the translations of five Hindi words were changed because the original translations involved several repetitions. The new forms were suggested by a native speaker who had no knowledge of the purpose of the experiment and was simply asked to provide an alternative translation in each case.

The variations of procedure used in the present experiment were first suggested by the following statement in which Brackbill and Little (1957, p. 314) explain their choice of the Same–Different procedure:

> A pilot study using an arbitrary list of 36 pairs of antonyms indicated that subjects did as well at guessing meanings when each English word was presented individually with either its foreign equivalent *or* opposite, as when the English and foreign antonyms were presented in pairs.

This is a result which ought not to occur if there is anything in the phonetic symbolism hypothesis

and which the present investigators judged would not occur with the original list. Therefore, it was decided to present that list again in the original manner, in the Same–Different manner, and with Foreign-Foreign pairings as used by Maltzman et al., for which they did not obtain significant results.

METHOD

For all presentations *Ss* received the same introduction describing the general idea of sound symbolism (no examples used) and instructing them to make their choices on the basis of the meanings suggested by the foreign words. Data from *Ss* who had any knowledge of either Chinese or Hindi were excluded from analysis. All presentations were audio-visual with the pronunciations of the foreign words recorded by native speakers.

English-Foreign-Pairs Presentation

This was as in the study of Brown et al. (1955). The English-Chinese list was first and the English-Hindi list second.

Foreign-Foreign Presentation

The *Ss* were presented with a Chinese pair and an equivalent Hindi pair (in random arrangement) and required to match correctly between languages. This differs from the procedure of Maltzman et al. (1956) only in presenting both pairs rather than a stimulus word and a response pair.

Same–Different Presentation

There were two lists presented in the order English-Chinese and then English-Hindi. For each position on each list a random event determined whether the pair would be a Same pair or a Different pair (with the restriction that there should be equal numbers of the two). For each foreign language pair a random event selected one word to be a Same and one to be a Different and a random event allotted the Sames and Differents to positions on the list. The English words assigned to the various Different foreign words were drawn at random from the English words remaining after the Same matches had been made. As in the study of Brackbill and Little (1957), therefore, Different pairs were not necessarily antonyms. Since this presentation deals with only two words at a time there were twice as many choices for each list as on the other two presentations (42 instead of 21). The *Ss* were told that half of the pairs were the Same.

Subjects

For the English-Foreign-Pairs and Same–Different presentations *Ss* were two large classes at Boston University, parallel sections in a course called Introduction to Literature. For the Foreign-Foreign presentation *Ss* were volunteers recruited at M.I.T. from classes in introductory psychology and convened as a group in a room of about the same size as those used at Boston University. Previous work on phonetic symbolism indicates that results do not depend on the age, sex, or education of *Ss* and so perfect comparability of the groups is not important. However, the two classes at Boston University, at least, were very much alike.

TABLE 1
PERCENTAGES OF CORRECT TRANSLATIONS FOR ENGLISH AND FOREIGN WORDS

Languages	Method of Presentation		
	English–Foreign Pairs ($N = 84$)	Foreign–Foreign ($N = 71$)	Same–Different ($N = 96$)
English-Chinese	60.2%***		51.7%**
English-Hindi	60.5%***		51.7%**
Chinese-Hindi		52.2%*	

* Significant at <.05 level.
** Significant at <.01 level.
*** Significant at <.0001 level.

RESULTS

The major results which are most comparable with those reported in the three studies that have been discussed are presented in Table 1. For all groups and lists judgments were significantly above chance. However, judgments with the English-Foreign-Pairs presentation were significantly better than the judgments with either of the other two presentations ($P < .0001$). The difference between the Foreign-Foreign and Same–Different presentations was not significant. Differences between lists within either the English-Foreign-Pairs or the Same–Different presentations were not significant.

DISCUSSION

The use of procedures like those used by Maltzman et al. (1956) and by Brackbill and Little (1957) has produced a set of results very much like those of the two earlier studies: percentages vary slightly above chance which are here significant because of the large *N*s but which with smaller groups would sometimes not be significant. These results were obtained with lists of words which in the English-Foreign-Pairs procedure of Brown et al. (1955) yield a very significant phonetic symbolism effect. Evidently the phenomenon is highly sensitive to the method of presentation.

What does this suggest about the nature of phonetic symbolism?

Sapir (1929) and Newman (1933) found with nonsense syllables that differences of magnitude are reliably suggested to English speaking *S*s by several phonetic contrasts. Back vowels, like *o* and *u*, seem larger than front vowels, like *i* and *e*; voiced consonants, like *b* and *d*, seem larger than voiceless consonants like *p* and *t*; velar articulations, like *g*, seem larger than bilabials, like *b*, and both of these seem larger than alveolars like *d*. Neither Sapir nor Newman varied the number of syllables in an invented word, but Wissemann (1954) has shown that a polysyllable suggests a more elaborately differentiated referent than does a monosyllable. Now let us suppose that *S* is in possession of these rules (without having formulated them). Only the English-Foreign-Pairs procedure, of the three procedures, should allow him to derive full advantage from his knowledge.

Suppose the English pair were *large–small*. As it happens this pair is not itself very good from the point of view of phonetic symbolism; *huge–tiny* is a great deal better. In one foreign language, antonyms naming the extremes of a dimension of magnitude might be appropriate by virtue primarily of vowel contrast; in another, by contrast of voiced and voiceless consonants; in another, by contrast of alveolar and velar consonants; in another, by contrast of the polysyllable with the monosyllable. If *S* were given the English pair he would know the referent dimension to be magnitude and so could correctly match any of the foreign pairs. However, if two foreign pairs were combined he might fail to identify the referent dimension and so match words on the basis of particular phonetic resemblances. These would not reliably produce the correct result since even in that most favorable case where both foreign pairs are good ones, the different languages may represent a given semantic contrast with different, though equally appropriate, phonetic contrasts. Of course it would sometimes happen, if both foreign pairs were good ones, that *S* would abstract the contrast in magnitude common to the two and match appropriately. Perhaps this is why the results with the procedure are slightly above chance. In many cases, however, only one of the two pairs will be appropriate to its referent and, in these circumstances, *S* would have small prospect of matching correctly.

Finally, there is the Same–Different procedure. One general difficulty here is that any single foreign word will not be either absolutely large or small or absolutely strong or weak or absolutely anything, and without its antonym or some contrast word in the same language *S* is not in a position to make the required relative judgments. Consider the Chinese *tŏ*. If it were matched with the English *one* *S*s might think of the *one–many* contrast and so have magnitude in mind. Now *tŏ* suggests a small magnitude when it is opposed to *go* but a large magnitude if opposed (as, in fact, it is in Chinese) to *yĭ*. Not knowing the contrast word in Chinese, *S* may again be expected to base his judgments on particular English-Chinese phonetic similarities and these may lead him to the right choice with slightly better than chance frequency but not with so high a frequency as when both foreign words are known. This example represents the most favorable case where an English word is matched with either its equivalent or contrary, but with the Same–Different procedure it must often happen that the English word is matched with a foreign word which is neither equivalent nor contrary but simply different. *S*'s task is likely to be still more difficult. Brackbill and Little (1957) have neatly demonstrated that their *S*s actually matched words on the basis of phonetic similarities in the pair presented. This behavior, which cannot very reliably lead to correct results, may have been forced by their method.

The present results suggest a new and more precise formulation of the phonetic symbolism hypothesis. Antonyms naming opposite ends of sensible referent continua (in languages with which *S* is unfamiliar) contrast with one another on appropriate phonetic dimensions often enough to make possible better than chance matches with equivalent terms in *S*'s native language. This condition, if it exists, need not be explained by a myth about the origin of speech in physiognomic representation. It is possible that in the history of languages antonyms evolve toward phonetic contrasts appropriate to their semantic contrast and that pairs so contrasting sound "right" to native speakers and so have a superior prospect for survival.

There remains some ambiguity in the notion of "sensible" continua as we are not sure exactly which referent dimensions to include. It is possible that the hypothesis should be restricted to continua concerned with magnitude and such common associates of magnitude as *heavy–light*, *slow–fast*, *dark–bright*, etc. The lists of Brown et al. (1955) and of Maltzman et al. (1956) contain many words of this kind and in the experiments of Sapir (1929) and Newman (1933) definite rules were discovered for the appropriate representation of these semantic dimensions. We mean next to test the phonetic symbolism hypothesis in this more specific form.

It should finally be noted that the results of the present experiment exclude that ever-present alternative explanation in work of this kind—ESP.

The three procedures should lead to the same degree of success if *S* were operating with telepathy or clairvoyance, but they, in fact, lead to quite different degrees of success.

SUMMARY

Using lists of English and foreign words three groups of *S*s were presented with the words in three different ways and in every case asked to match words likely to have the same meaning. Matching was correct at levels significantly above chance for all lists and all procedures but very significantly superior results were obtained when *S* was presented with an English pair and a foreign pair than when both pairs were foreign or when only one English and one foreign word were presented and had to be judged the same or different. It is concluded that the phonetic symbolism effect is sensitive to changes of procedure in a way that suggests that phonetic symbolism in natural languages may be limited to pairs naming the extremes of sensible continua and, conceivably, it may be limited to pairs naming continua concerned with magnitude and its concomitants.

REFERENCES

BRACKBILL, YVONNE, & LITTLE, K. B. Factors determining the guessing of meanings of foreign words, *J. of abnorm. soc. Psychol.*, 1957, **54**, 312–318.

BROWN, R. W., BLACK, A. H., & HOROWITZ, A. E. Phonetic symbolism in natural languages, *J. abnorm. soc. Psychol.*, 1955, **50**, 388–393.

MALTZMAN, I., MORRISETT, L., & BROOKS, L. O. An investigation of phonetic symbolism, *J. abnorm. soc. Psychol.*, 1956, **53**, 249–251.

NEWMAN, S. S. Further experiments in phonetic symbolism, *Amer. J. Psychol.*, 1933, **45**, 53–75.

SAPIR, E. A study in phonetic symbolism, *J. exp. Psychol.*, 1929, **12**, 225–239.

WISSEMANN, H. *Untersuchungen zur Onomatopoiie*, Heidelberg: Carl Winter, Universitätsverlag, 1954.

(Early publication received March 12, 1959)

LOOKING FOR A BOOK PUBLISHER?

It is no secret that publication is regarded as the foundation stone of a scholar's career. In order to gain recognition, prestige, and advancement, the scholar must seek publication.

Exposition Press offers scholars a complete publishing service, under our special academic imprint, *Exposition-University Books.* Though the proportion of scholarly publications in the lists of the trade houses is shrinking, due to economic pressures, several of our recent titles in the academic fields have gained popular reviews, professional recognition and steady sales. These books are the result of a publishing plan that is bringing a steady flow of scholarly writers to Exposition Press. They reflect the high editorial standards and quality of design and production which have won the respect of libraries, schools, booksellers and critics.

FREE! The behind-the-scenes story of book publishing revealed in two fact-filled, illustrated brochures, containing a detailed description of our 40%-royalty subsidy plan including a breakdown of contract terms and typical costs (in print for the first time). Copies are available on request. Your inquiries and manuscripts are invited. An editorial appraisal will be furnished promptly without obligation.

Please write to **Jonathan Phipps,** *Editorial Department.*

EXPOSITION PRESS, *Main Office, 386 Park Ave. S., New York 16, N.Y.*

SYMPOSIUM: PSYCHOANALYSIS AS SEEN BY ANALYZED PSYCHOLOGISTS

This well-known symposium, which first appeared in the *Journal of Abnormal and Social Psychology* in 1940, has now been reprinted. The contributors are Edwin G. Boring, Carney Landis, J. F. Brown, Raymond R. Willoughby, Percival M. Symonds, Henry A. Murray, Else Frenkel-Brunswik, David Shakow, Donald V. McGranahan, and Austin B. Wood. Comments on the symposium are made by Hanns Sachs and Franz Alexander. 160 pages.

Price, $2.00

Order from

American Psychological Association

1333 Sixteenth Street N. W. **Washington 6, D. C.**

Recent developments, therapeutic advances in . . .

SCHIZOPHRENIA

An Integrated Approach

Edited by ALFRED AUERBACK, M.D.

with 16 Contributors

Sponsored by the American Psychiatric Association, this important new book surveys recent progress in the understanding and treatment of schizophrenia. Presenting the most significant portions of the scientific data available, it covers the latest psychotherapeutic techniques. Reviews Russian developments in neurophysiology; outlines current biochemical studies on the psychotoxic blood fraction, taraxein; and appraises the narcoleptic drugs used in treatment. *Illus., 22c pp.* $5.50

ADOLESCENT AGGRESSION

ALBERT BANDURA, Stanford University, and RICHARD H. WALTERS, University of Toronto

Just published! This revealing study examines the influence of child-training practices and family interrelationships on the development of antisocial, aggressive behavior in adolescent boys. Based on interviews and projective tests of boys and their parents, the study employs a carefully matched control group. Extensive quotations from the interviews allow the subjects to speak for themselves. *A volume in a Series on Psychology under the general editorship of Dr. J. McV. Hunt, University of Illinois.*
72 illu., tables; 330 pp. $7.50

THE PSYCHOLOGY OF EXCEPTIONAL CHILDREN

KARL C. GARRISON, University of Georgia, and DEWEY G. FORCE, Jr., University of Minnesota

Third Edition of this well-known book discusses the nature and identification of exceptional children and the ways in which exceptionality affects emotional, social, and educational adjustment. The book summarizes current developments, concepts, and research; includes new material on such areas of exceptionality as epilepsy, cardiac conditions, and cerebral palsy. "*. . . a sound treatment of the psychology of exceptional children.*"—A. M. Jordan, University of North Carolina. *103 illu., tables; 526 pp.* $6.00

THE ABNORMAL PERSONALITY

ROBERT W. WHITE, Harvard University

This authoritative book presents the study of disordered personalities as it relates to the larger understanding of human nature. Examines maladjustment and neurosis; discusses normal development and shows how it can go astray. Covers psychotherapy, delinquency, brain injuries, etc. "*Dr. White's book is without question the best available text on abnormal psychology today.*"—Peter Madison, Swarthmore College. *2nd Ed. Illus., 604 pp.* $7.00

Order your books direct from:

THE RONALD PRESS COMPANY
15 East 26th Street, New York 10, N. Y.

The Natural History of Cerebral Palsy

By Bronson Crothers and Richmond S. Paine. The emphasis in this volume is upon the effect of fixed lesions arising early in life upon growth and development. Essentially the problems considered here are those which arise as the handicapped child faces the stresses of full competition. The authors report on a large series of cases studied over a period of years, and endeavor to appraise the validity of treatment given.
$6.75

Psychological Appraisal of Children with Cerebral Defects

By Edith Meyer Taylor. Neurologically impaired children have too often been treated as if all were alike, or all different, and too little attention has been paid to the fact that the same child, with the same problem, may appear very different at various ages. This long-needed, practical guide details processes of evaluation that have proved most useful, and places emphasis upon intellectual and socio-emotional development of the child. Illustrated. A Commonwealth Fund Book
$6.50

Longitudinal Studies of Child Personality

ABSTRACTS WITH INDEX

By Alan A. Stone and Gloria Cochrane Onqúe. This thorough survey provides an analytical bibliography summarizing all important research through 1965. The index is divided by subject matter and is a complete reference to the 207 entries in the abstract section. A Commonwealth Fund Book
$5.00

Teaching Comprehensive Medical Care

A PSYCHOLOGICAL STUDY OF A CHANGE IN MEDICAL EDUCATION

By Kenneth R. Hammond, Fred Kern, Jr., and others. With a foreword by Ward Darley, M.D. A comparative study and evaluation of two kinds of medical education conducted at the University of Colorado by a research team including four psychologists. Existing psychological procedures were adapted, and interesting new methods developed.
$8.00

Through your bookseller, or from

HARVARD
University Press
79 Garden Street, Cambridge 38, Mass.